

## 基于粗糙集理论的地矿信息量化分析

朱雅琼,袁艳斌,周 尤,彭晶倩

(武汉理工大学资源与环境工程学院,武汉 430070)

**摘要:** 在地学研究中占主导地位的是高维非线性的复杂问题,粗糙集理论对分析此类问题有着明显的优势。研究粗糙集在地矿信息挖掘中的应用,能更加充分地利用地矿勘查数据,揭示数据背后隐藏的规则和联系,成功地选择特征矿化标志,建立基于粗糙集理论的矿化信息研究模型,定量描述采样点性质,为靶区预测提供科学的依据。

**关键词:** 粗糙集; 地矿信息; 信息系统

中图分类号: G 202

文献标识码: A

文章编号:1671-4431(2008)01-0123-03

### The Quantitative Analysis of Geological and Mineral Information Based on Rough Set Theory

ZHU Ya-qiong, YUAN Yan-bin, ZHOU You, PENG Jing-qian

(School of Resources and Environmental Engineering, Wuhan University of Technology, Wuhan 430070, China)

**Abstract:** Geological research was dominated by complicated NP (Non-deterministic Polynomial) problem. It's dominant for Rough Sets to analyze this kind of problems. Researches on the application in the geological and mineral information mining could make full use of the data from geological and mineral exploration, reveal the hidden rules and relation behind the data, choose the mineralization genes successfully, establish the study model of mineralization information based on Rough Sets, describe the character of sample points in quantification and then offer the scientific basis for target prediction.

**Key words:** rough set; geological and mineral information; information system

地矿数据具有多来源、多维数、多类别、多变量和多应用主题的“五多”特征<sup>[1]</sup>,在地学研究中占主导地位的是高维非线性的复杂问题<sup>[2]</sup>,从海量的地矿勘测数据中提取特征矿化信息是当前定量化矿产勘查重要研究课题之一。粗糙集能直接从给定问题的描述集合出发,通过不可分辨关系和不可分辨类确定近似域,从而找出问题的内在规律<sup>[3]</sup>。该方法不需数据的附加信息或先验知识,在信息不丢失、知识库分类能力不变的前提下,删除无关或不重要的属性,得到的知识以规则的形式描述<sup>[4]</sup>。粗糙集完全适用于矿产勘查“灰色”的特点<sup>[5]</sup>,基于该理论分析地矿信息是一种新的探索。

### 1 粗糙集理论

系统的形式为 $(U, A, F, d)$ ,其中, $U$ 为论域, $U$ 中的元素 $x_i (i \leq n)$ 称为研究对象,记论域为 $U = \{x_1, x_2, \dots, x_n\}$ ,描述研究的对象全体; $A$ 为每个研究对象对应的属性集,记为 $A = \{a_1, a_2, \dots, a_m\}$ , $A$ 中的每个元素 $a_l (l \leq m)$ 描述一个属性; $F$ 为 $U$ 与 $A$ 之间的关系集,即 $F = \{f_l: U \rightarrow V_l (l \leq m)\}$ ,其中 $V_l$ 为 $a_l (l \leq m)$ 的值域; $d$ 为决策信息, $U \rightarrow V_d, V_d$ 取有限值。

收稿日期:2007-09-20.

基金项目:国家自然科学基金(40572166,50309013)和湖北省自然科学基金(2005ABA228).

作者简介:朱雅琼(1983-),女,硕士生.E-mail:zhuyq06@126.com

粗糙集通过寻找属性核去掉多余的属性,约简决策表,从而获取相应的确定规则。如果2个地质工程点所获取的属性信息完全相同,而决策属性值不同,表明数据库中一些属性丢失<sup>[6]</sup>,粗糙集不移除这些不协调的数据,而是直接从原始决策表中求取近似集,并运用推理引擎,分别从下近似集中获取确定规则,从上近似集中获取可能规则,使问题描述更加精确。基于粗糙集的规则获取的2种模式如图1所示<sup>[7]</sup>。

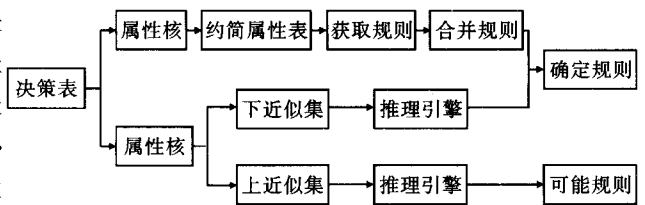


图1 基于粗糙集规则的推理模式

## 2 基于粗糙理论的地矿信息分析

### 2.1 研究流程

图2为工作流程图,研究步骤如下:1)选择地质变量,建立标志总体。从研究区大量的图件和表格中提取与矿化可能有关的各种地质因素,组成数据集。2)基于粗糙集的数据预处理。对于非数值属性,即名义型数据进行编码,数值化数据集,对连续型属性值选择合适的方法离散化,形成决策表。3)属性约简。直接从数据上进行分析,删除冗余属性和不必要属性,求出属性核。4)属性值约简。删除冗余属性值,提出规则知识。5)在4)的基础上,建立靶区预测模型

$$P = f(x_1, x_2, \dots, x_n) = \sum_{i=1}^n a_i x_i$$

其中, $P$ 为矿床值的变量; $a_i$ 为各控矿标志权值(信息权),其大小表示该标志的重要性程度; $x_i$ 为各控矿因素分权,区分哪些标志起正作用,哪些标志起负作用,以评价各标志对研究对象所起作用的性质及大小。6)解释评价。参照已有的地质理论或假说,对各地质标志的意义根据计算结果重新估价。

### 2.2 实例分析

#### 2.2.1 决策表的构建

提取该研究区关于地层、断裂、褶皱、岩性、围岩蚀变等地质特征,其描述见表1。分析研究区1:2 000地质图以及剖面图,采样点根据钻孔、探槽、浅井等提供的信息描述,组成数据集。对于名义型数据,如断层、岩性等,

地质特征	勘查特征描述
地层	下泥盆统坡松冲组(一段 $D_1ps^1$ 、二段 $D_1ps^2$ ),下奥陶统闪片山组( $O_1s$ )
断裂	近南北向断裂,近东西向断裂,北西向断裂
褶皱	背斜
岩性	辉绿岩、石英砂岩、粉砂质泥岩、石英砂质粘土、灰岩
围岩蚀变	硅化、褐铁矿化、黄铁矿化、辉锑矿化、碳酸盐化、粘土化、绢云母化等

研究采用“0-1”二态数据描述采样点某种属性的有或无;对于连续属性值,如采样点的金品位,研究粗糙集理论和多智能体技术对矿产预测数据的离散化方法。

表2 形成决策表

$U$	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$a_6$	$a_7$	$a_8$	$a_9$	$a_{10}$	$a_{11}$	$a_{12}$	$a_{13}$	$a_{14}$	$d$
$x_1$	1	0	0	1	1	1	1	0	0	0	1	0	0	0	1
$x_2$	1	0	0	1	1	0	1	0	0	0	1	0	0	0	1
$x_3$	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0
$x_4$	0	1	0	0	1	0	0	1	0	1	0	0	0	0	0
$x_5$	1	0	0	1	1	0	1	0	1	0	1	1	0	0	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

表2即为部分采样点信息。其中,包括20个研究对象,14种属性, $a_{ij}(i \leq 20, j \leq 14)$ 均为属性离散值,

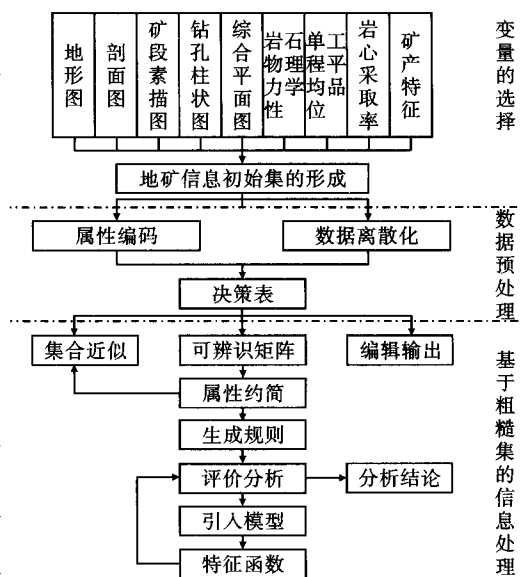


图2 工作流程图

分别用1、0表示属性的有、无; $d_i (i \leq 20)$ 为决策属性,以地质工作者圈定研究区矿体边界值为依据,将金品位分为高、低2个等级,分别用整数1、0表示。

其中  $a_1, a_2, a_3$  分别表示地层  $D_{1ps}^1, D_{1ps}^2, O_1s$ ;  $a_4, a_5$  分别表示断裂、背斜;  $a_6, a_7, a_8, a_9, a_{10}$  分别表示辉绿岩、石英砂岩、粉砂质泥岩、石英砂质粘土、灰岩;  $a_{11}, a_{12}, a_{13}, a_{14}$  分别表示硅化、粘土化、褐铁矿化、黄铁矿化。

### 2.2.2 属性约简和规则的获取

基于可识别矩阵提取决策表的属性核,找到约简属性集<sup>[8]</sup>。取集合  $O$  代表金品位高的研究对象的集合,记  $O = \{x_h | h \leq n\}$ ;取  $F$  为论域  $U$  与属性  $A$  之间的关系集,根据属性划分等效关系,记  $R = \{(x_i, x_j) | f_i(x_i) = f_i(x_j)\}$ ,则  $R$  是  $U$  上得等价关系,  $[x]_R$  表示包含  $x$  的等价类。若同一等效类中有2个或者2个以上的对象,则这些对象不可分辨。通过集合  $O$  与不可分辨对象的比较分析,可得到下逼近区和上逼近区。下近似集:  $POS(O) = \min\{[x]_R \subseteq O\}$ ;上近似集:  $POS(O) + BAN(O) = \max\{[x]_R \subseteq O\}$ 。从通过以上分析,获得表3中的有用规则。

### 2.2.3 成果分析

金品位高的研究对象特点:1)地层:主要位于  $D_{1ps}^1$ ;2)构造:断层部位样点金品位高;3)岩性:石英砂岩;4)蚀变:主要有硅化、粘土化。结论符合研究区的实际情况,该研究区金矿产于那洒破背斜北翼或倾伏端、加里东不整合面附近的硅化岩带及其与断裂带重合叠加部位;目前发现的金矿体均赋存于坡松冲组底部( $D_{1ps}^1$ )地层中;含矿岩性主要是硅化石英砂岩,局部辉绿岩;金矿化富集地段以硅化、粘土化为主,多种蚀变叠加时,金矿化较强。试验证明,采用粗糙集这一数学方法,可以成功提取特征矿化信息。

## 3 结 语

矿产资源评价的过程就是信息的搜集、整理、处理、成矿信息的提取、综合分析、成矿区带或找矿靶区的确定以及成果表示的过程。在进行信息合成时,只有采用一定的数据模型对各类无直观规律的数据集进行整理、分析,才能进而进行不同种类的信息集成工作。在此基础上深层次的研究基于粗糙集理论的定量机制,建立定量的成因和空间分布数学模型,建立特征函数,查明各控矿因素和找矿标志的找矿信息量,如各因素标志的信息量或信息权,各因素标志的找矿概率,各因素标志最有利成矿或找矿的数值区间等,直接从数值上评定矿点性质,为靶区预测提供科学的依据。

### 参考文献

- [1] 袁艳斌,汪新庆,刘刚. 三峡坝区工程地质信息系统集成开发研究[J]. 地球科学-中国地质大学学报, 1999, 24(5): 542-544.
- [2] 李双成,郑度. 神经网络模型在地质研究中的应用进展[J]. 地球科学进展, 2003, 18(1): 68-76.
- [3] 瞿彬彬,卢炎生. 基于粗糙集的不完备信息系统规则推理算法[J]. 小型微型计算机系统, 2006, 27(4): 698-701.
- [4] 王霞,唐德善. 基于粗糙模糊集的区域水资源系统的评价方法[J]. 水利规划与设计, 2006(1): 31-33.
- [5] 赵洁心,冯波,谭俊,等. 我国矿产资源开发利用现状与可持续发展探讨[J]. 黄金, 2006, 5(27): 1-4.
- [6] Guo Jiayuan. Rough Set-based Approach to Rule Generation and Rule Induction[J]. International Journal of General Systems, 2002, 31(6): 601-617.
- [7] Pawlack Z. Rough Sets[J]. Communications of the ACM, 1995, 38(11): 89-95.
- [8] 王国胤. 粗糙集理论与知识获取[M]. 西安:西安交通大学出版社, 2001.

表3 获取的规则

规则	可信度
$a_1(1) \Rightarrow d(1)$	1.00
$a_4(1) \Rightarrow d(1)$	1.00
$a_7(1) \Rightarrow d(1)$	1.00
$a_{11}(1) \Rightarrow d(1)$	0.83
$a_7(1) \text{ AND } a_9(1) \Rightarrow d(1)$	0.67
$a_1(1) \text{ AND } a_9(1) \Rightarrow d(1)$	0.67
$a_{11}(0) \text{ AND } a_{12}(1) \Rightarrow d(0)$	1.00
$a_{12}(1) \Rightarrow d(1)$	0.67
...	...