

文章编号 1001-8166(2008)02-0214-05

金矿资源定量预测的粗糙集方法^{*}

朱雅琼 袁艳斌 周 尤 彭晶倩 詹云军
(武汉理工大学资源与环境工程学院,湖北 武汉 430070)

摘要 :矿产信息是各种成矿相关信息的综合体现,为了有效地提取成矿预测综合信息,有必要客观地筛选原始观测信息,突出成矿密切相关的致矿因子。粗糙集不需要数据的附加信息或先验知识,在知识库分类能力不变的前提条件下,删除无关或不重要的属性,能对决策系统进行有效约简。提出基于粗糙集理论进行集成化预测模型研究的新方法,基于粗糙集思想提取与成矿密切相关的特征矿化信息,获取最佳变量组合及区间值,并将其作为参量建立预测模型,结合经典矿床统计预测聚类方法确定代判临界值,及特征分析方法对矿产资源进行定量预测,确立了8个成矿有利单元,与研究区勘查工程资料基本吻合,表明该方法能够有效降噪,简化模型,为靶区预测提供准确的依据。

关键词 :矿产资源;粗糙集;定量预测;特征分析

中图分类号 :P618.51;P624.7 **文献标识码** :A

1 引言

矿产信息是各种成矿相关信息的综合体现,包括地质、地球化学、地球物理以及由它们伴生的多种矿化信息,当代计算机科学和数学地质方法为找矿靶区优选和综合评价提供了技术和方法支持,对成矿信息模型进行全面探索、对多元的地质信息定量分析是矿产勘查和成矿预测评价的发展方向之一。

目前已有的单一确定型或随机型数学模型,在表达复杂的地质历史过程中,由于条件的限制,都存在各自的局限,这也是数学地质目前尚不能很好解决地质学定量研究的根本症结。为了有效地提取成矿预测综合信息,有必要客观地对诸多原始观测信息进行筛选,突出成矿密切相关的致矿因子,得到最佳变量组合。我们在对诸多非线性数学方法的比较研究中,发现粗糙集^[1]不需数据的附加信息或先验

知识,在知识库分类能力不变的前提条件下,删除无关或不重要的属性,能对决策系统进行有效约简^[2~4]。在地质研究中占主导地位的是高维非线性的复杂问题^[5~6],针对该问题,基于粗糙集思想,对众多初始地质变量进行约简,用最佳变量组合参与建模,分析其成矿预测效果,发现约简变量建模比全局变量建模的拟合误差小,表明降维后的地质变量具有更好的预测效果,兼有降噪作用。

2 研究区概况及单元划分

2.1 研究区概况

本研究区位于华南褶皱系滇东南褶皱带文山—富宁断褶束西畴拱凹北缘,文山—那洒弧形构造东段。地壳活动经历了由地槽(加里东期)到地台(华力西期),再到地槽(印支期)的复杂演化过程。区域构造为环绕越北古陆呈同心环状展布的弧形构造,北西向断裂穿插其间,构成了本区构造格架。弧

* 收稿日期 2007-06-14;修回日期 2007-12-18。

* 基金项目:国家自然科学基金项目“粗糙集支持下特征矿化信息挖掘的粒子群演化方法”(编号:40572166)和“基于混沌进化算法的数字流域信息挖掘与复合”(编号:50309013);湖北省自然科学基金项目“数字流域空间信息融合机理研究”(编号:2005ABA228)资助。

作者简介:朱雅琼(1983-),女,湖北武汉人,硕士生,主要从事信息科学在资源环境中的应用研究。E-mail:zhuyq06@126.com

形构造由一系列弧形弯曲的断裂和线状褶皱组成 ,弧顶向北突出 ,呈半环状展布 构造变形主要为晚印支运动产物。区域矿产丰富 ,现已发现的矿(床)点有九克、皂角树、革夺、韭菜坪、俄里、田坊等(锑)金矿床(点) ,老寨湾金矿是 1991 年发现的。

2.2 金矿地质单元

老寨湾成矿地段地质勘探程度较高 ,已经发现一些矿床和矿点 ,为了进一步挖掘找矿潜力 ,首先对该矿区进行规则网格划分单元 ,编号单元为已勘探单元(图 1)。其中勘探程度高的 3 号、4 号、6 号、7 号、8 号、11 号、12 号、13 号、18 号、19 号、20 号、22 号、23 号、24 号、28 号、30 号、31 号和 35 号单元为模型单元 ,建模后预测剩余单元情况。

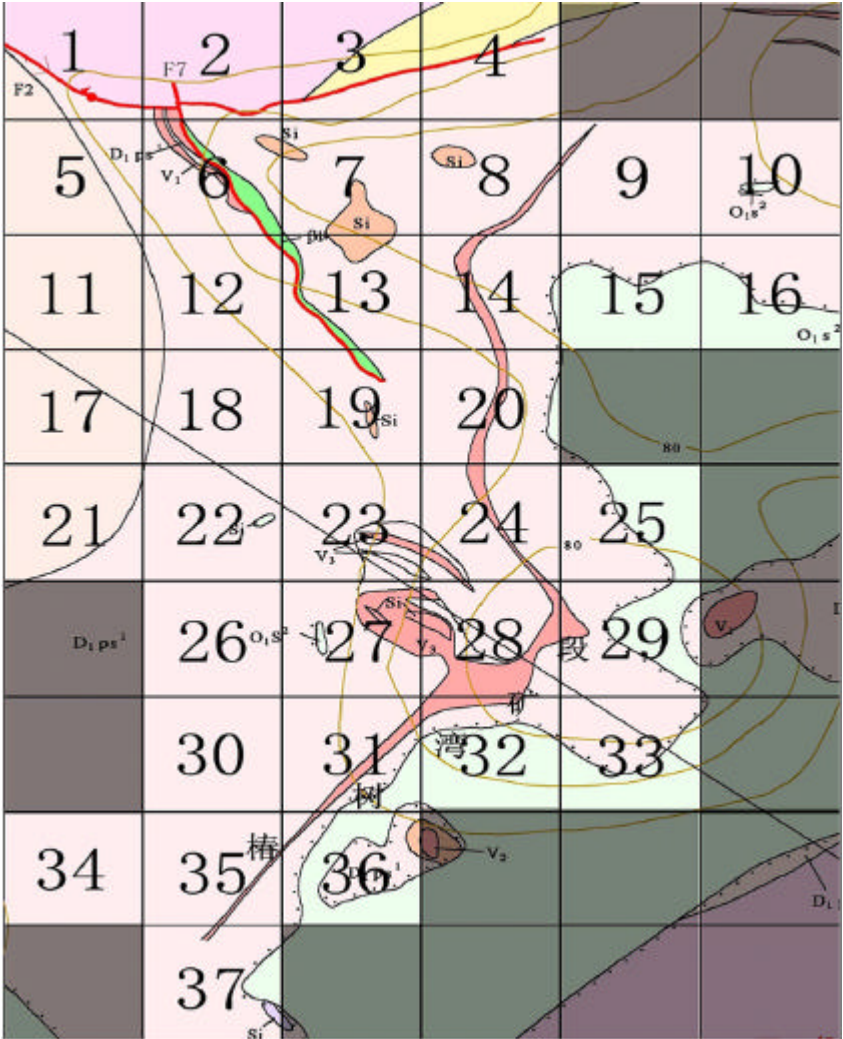


图 1 老寨湾矿区网格划分单元示意图
Fig.1 Diagram of grid cells of research area

$$M(i,j)=\begin{cases} \{a_k \mid a_k \in B \wedge a_k(x_i) = a_k(x_j)\}, & d(x_i) \neq d(x_j) \\ d(x_i) = d(x_j) \end{cases} \quad (1)$$

如果矩阵中某个元素只有一个属性 a_k ,则表明该属性是区分这个矩阵元素所对应的两个样本所必须的属性 ,也是唯一能够区分这两个样本的属性 ,即可辨识矩阵中的这些元素所包含的属性组成的属性集合就是该决策表系统的相对属性核。

4 研究流程

- (1) 选择地质变量 ,建立找矿标志总体。提取研究区各种地质异常 ,组成初始数据集。
- (2) 基于粗糙集方法的数据预处理。对于非数值属性进行编码 ,数值化数据集 ,对连续型属性值选

3 粗糙集理论

问题研究中将信息系统中的属性分为条件属性和决策属性两类 ,针对一些研究对象 ,如各类地质工程点等 ,提取特征矿化信息作为矿产资源定量研究和定量预测的依据。

3.1 信息系统

我们把该区矿产资源预测问题作为一个约束系统来加以研究 ,其系统的形式可表示为 (U, A, F, d) ,其中 U 为论域 , U 中的元素 $x_i (i = 1, \dots, n)$ 称为研究对象 ,一般记论域为 $U = \{x_1, x_2, \dots, x_n\}$,描述研究的对象全体 ;每个研究对象对应的属性集为 A ,记为 $A = \{a_1, a_2, \dots, a_m\}$, A 中的每个元素 $a_i (i = 1, \dots, m)$ 描述相应的一个属性 ; F 为 U 与 A 之间的关系集 ,即 $F = \{f_i : U \rightarrow V_i (i = 1, \dots, m)\}$,其中 V_i 为 $a_i (i = 1, \dots, m)$ 的值域 , d 为决策信息 , $U \rightarrow V_d, V_d$ 取有限值。

3.2 不可分辨关系

属性集 A 的任意子集 B 定义论域 U 上的一个二元关系 $R_B, x_i, x_j \in R_B$ 。给定属性集 B ,如果对于任意 $a \in B, a(x_i) = a(x_j)$,则 x_i 和 x_j 不可分辨 ,其中 $a(x)$ 表示元素 x 对于属性 a 的属性值 ,记为 $R_B = \{(x_i, x_j) \mid f_i(x_i) = f_i(x_j) (a_i \in B)\}$ 。 R_B 表示等价关系 , R_B 所有等价类的集合记为 U/R_B 。如果 (x_i, x_j) 属于关系 R_B ,那么 x_i, x_j 记为 B -不可分辨 ,关系 R_B 的等价关系定义为 B -基本集。

3.3 属性约简

在粗糙集理论中 ,属性约简基于不分明矩阵。 B 是条件属性集 , D 是决策属性集 , $a_i(x_j)$ 表示对象 x_j 在属性 a_i 上的属性值 , $M(i, j)$ 表示不分明矩阵中第 i 行 j 列的元素 ,则可辨识矩阵表示为 :

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 ,删除冗余属性和非特征属性 ,求出属性核。

(4) 建立靶区预测模型。

$$P = f(x_1, x_2, \dots, x_n) = \sum_{i=1}^n a_i x_i \quad (2)$$

其中 : P 为关联度 ; a_i 为各控矿标志权值 ; x_i 为各控矿因素分权。

(5) 预测。基于最优变量组合构建特征函数 ,对地质单元进行成矿预测。

5 老寨湾金矿区成矿有利地段定量预测

5.1 变量提取和赋值

对研究区勘查的数据进行整理 ,分析研究区观测数据和各大比例尺地质图和截面图 ,提取该研究区关于地层、断层、岩性、围岩蚀变等地质特征 ,其描述见表 1。

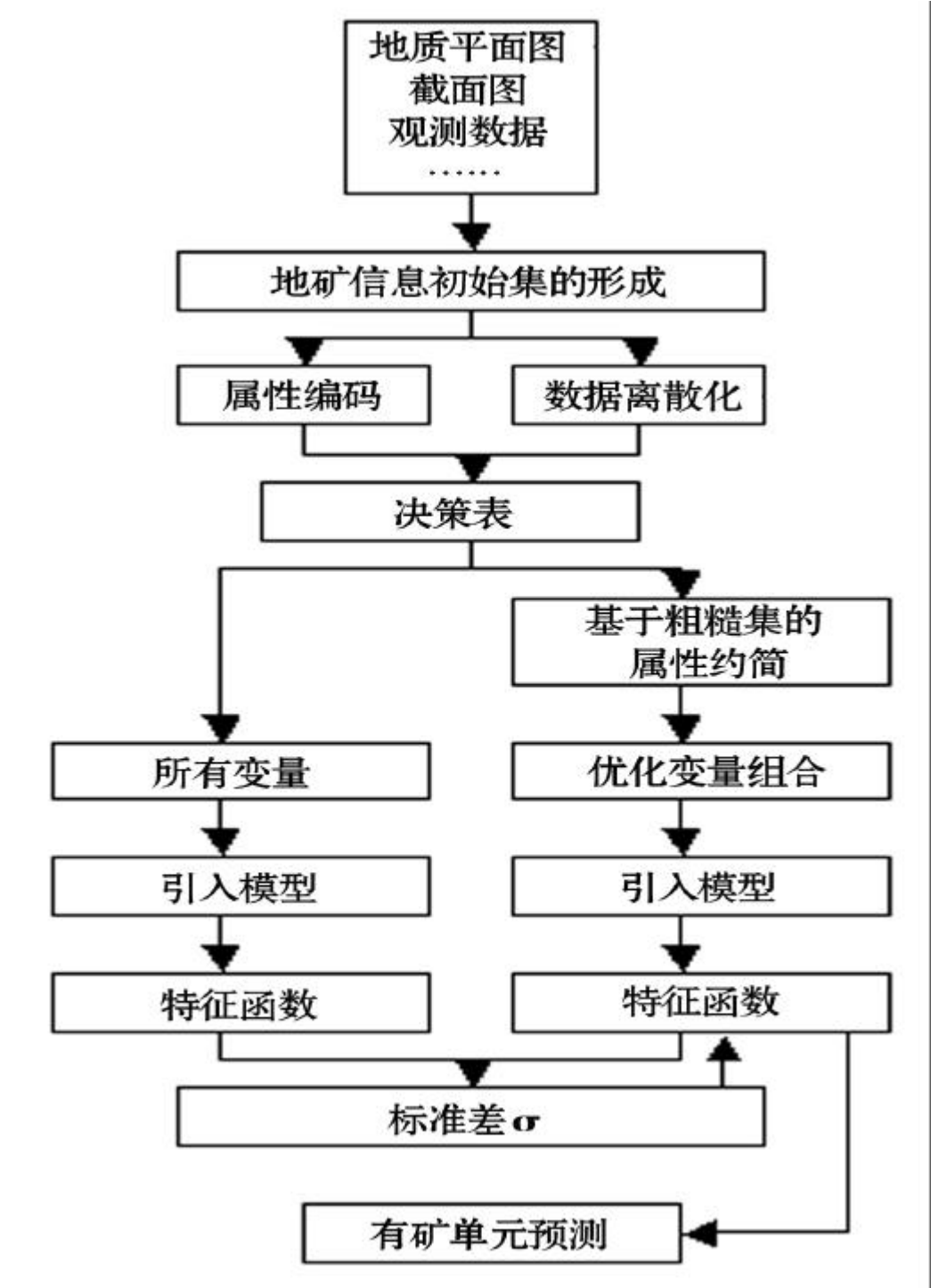


图 2 工作流程图

Fig.2 W ork flow

表 1 某矿区地质特征属性表

Table1 The attribute table of geological character in certain diggings	
地质特征	地质特征描述
地层	下泥盆统坡松冲组(不整合 ,一段 D_{1ps}^1 、二段 D_{1ps}^2)、下奥陶统闪片山组(O_{1s})
断层	近南北向断层 ,近东西向断层 ,北西向断层
岩性	辉绿岩、石英砂岩、粉砂质泥岩、灰岩
围岩蚀变	硅化、粘土化、褐铁矿化、黄铁矿化、辉锑矿化、碳酸盐化、绢云母化等

决策表的构建包括名义型数据的编码和多源勘查数据的离散化。在矿产信息中 ,名义型数据即一些描述性的属性值 ,如断层、岩性等 ,本研究中采用“二态数据” ,即“0 - 1”数据进行基于粗糙集数学分析的探讨 ,描述地质单元某种属性的有或无 ;金品位

需要进行离散化处理 ,将金品位分为高、低和零三类 ,分别用 2 ,1 和 0 表示。根据研究区经济技术条件和金矿成因类型 ,同时参照行业标准 ,断点选取为 0 ,0.5 和 1.0。如果金品位值在区间[0 ,0.5) 内 ,决策值为 0 ,如果金品位值在区间[0.5 ,1.0) 内 ,决策值为 1 ;在采样点中 ,金品位最大值为 1.84 g /t ,如果金品位值在区间[1.0 ,1.84] 内 ,决策值为 2。决策表见表 2。

表 2 决策表
Table 2 Decision table

地质单元	a ₁	a ₂	a ₃	a ₄	a ₅	a ₆	a ₇	a ₈	a ₉	d
3	1	1	0	1	1	0	1	1	0	1
4	1	1	0	1	1	0	1	1	0	1
6	0	1	0	1	0	1	1	1	0	2
7	0	0	0	1	1	0	0	1	0	0
8	1	0	0	1	0	0	1	1	0	2
11	0	0	0	1	1	1	1	0	0	0
12	1	1	0	1	0	1	0	0	1	2
13	1	1	0	1	1	0	1	0	1	2
18	1	0	0	1	0	0	0	0	0	0
19	1	1	1	1	1	1	1	1	1	2
20	1	0	0	1	1	0	1	0	1	2
22	1	0	0	1	1	1	0	0	0	1
23	1	0	0	1	1	1	0	0	1	2
24	1	0	1	1	0	0	1	0	0	2
28	1	0	0	1	1	0	1	1	1	2
30	1	0	0	1	1	1	0	0	0	0
31	1	0	0	1	1	1	0	0	0	1
35	1	0	0	1	0	1	0	0	0	0

a₁ :不整合 ;a₂ :断裂 ;a₃ :辉绿岩 ;a₄ :石英砂岩 ;a₅ :泥岩 ;a₆ :灰岩 ;a₇ :硅化 ;a₈ :粘土化 ;a₉ :褐铁矿化

5.2 特征地矿信息的提取

构建好决策表后 ,利用粗糙集对属性值进行约简 ,找出各矿化信息与成矿的关联度。基于粗糙集方法 ,从数量众多的初始变量中筛选最重要的变量 ,目的是要达到“变量结构最优化” ,即要具有最优变量组合。这种筛选可以减少空间维数 ,简化系统 ,同时又不损失与研究对象有直接和间接联系的主要信息。根据粗糙集的分析 ,可得属性核为{ a₅ } 和{ a₉ } ,属性约简集为{ a₁ , a₅ , a₇ , a₉ }。

5.3 成矿预测

(1) 基于最优变量建模。基于粗糙集方法 ,总结出综合找矿标志{ a₁ , a₅ , a₇ , a₉ } ,使用特征分析方法^[7 8]进行定位及定量预测 ,以本研究区勘测程度高的单元作为成矿预测的模型单元 ,采用特征分析中的乘积矩阵主分量法 ,确定相应各变量的变量权 ,进而建立起本研究区典型矿床的特征分析模型

如下：

$$Y =0.6338 \times a_1 +0.5570 \times a_5 +0.4509 \times a_7 +0.2911 \times a_9$$

(3)

(2) 基于所有变量建模。若所有变量参与建模 ,采用同样的方法确定相应变量的变量权 ,进而建立起本研究区典型矿床的特征分析模型如下：

$$Y =0.4665 \times a_1 +0.2150 \times a_2 +0.0710 \times a_3 +0.5399 \times a_4 +0.3934 \times a_5 +0.2693 \times a_6 +0.3338 \times a_7 +0.2329 \times a_8 +0.2160 \times a_9$$

(4)

(3) 比较分析。对于这两个模型 ,分别将研究区模型单元的地质标志取值逐一带入(3)式和(4)式 ,即得到各单元的关联度 Y 值。为了便于比较分析 ,将关联度和实际勘查品位值归一化处理 ,分别表示成矿概率估计值和成矿概率参照值：

$$P =\frac{Y_i - Y_{\min }}{Y_{\max } - Y_{\min }},i =1,2, \ldots ,n$$

计算结果见表 3。

估计值与参照值之间的误差采用公式：

$$= \sqrt{\frac{1}{n} \sum_{i=1}^n \left(x_{\text{估}} - x_{\text{测}} \right)^2}$$

最优组合变量参与建模 ,得 =0.3959 ;全部变量参与建模 ,得 =0.3986。图 3 为估计值与参照值对照图 ,其中估计值 1 为约简变量参与建模预测结果 ,估计值 2 为所有变量参与建模预测结果 ,参照值为研究单元勘查值 ,估计值与参照值之间的误差

表现为 < ,表明用约简后的组合变量{ a₁ , a₅ , a₇ , a₉}可以取代众多变量对研究区进行预测。

表 3 估计值与参考值对照表

Table 3 Comparison table of estimate value and reference value

地质单元	模型(a)		模型(b)		勘查值	
	关联度	归一化(%)	关联度	归一化(%)	金品位	归一化(%)
3	1.6417	80.36	2.1815	67.87	1	50.00
4	1.6417	80.36	2.1815	67.87	1	50.00
6	0.4509	0.00	1.5909	33.76	2	100.00
7	0.557	7.16	1.1662	9.23	0	0.00
8	1.0847	42.77	1.5731	32.73	2	100.00
11	1.0079	37.59	1.5364	30.61	0	0.00
12	0.9249	31.99	1.7067	40.45	2	100.00
13	1.9328	100.00	2.1646	66.89	2	100.00
18	0.6338	12.34	1.0064	0.00	0	0.00
19	1.9328	100.00	2.7378	100.00	2	100.00
20	1.9328	100.00	1.9496	54.48	2	100.00
22	1.1908	49.93	1.6691	38.28	1	50.00
23	1.4819	69.57	1.8851	50.75	2	100.00
24	1.0847	42.77	1.4112	23.38	2	100.00
28	1.9328	100.00	2.1825	67.93	2	100.00
30	1.1908	49.93	1.6691	38.28	0	0.00
31	1.1908	49.93	1.6691	38.28	1	50.00
35	0.6338	12.34	1.2757	15.55	0	0.00

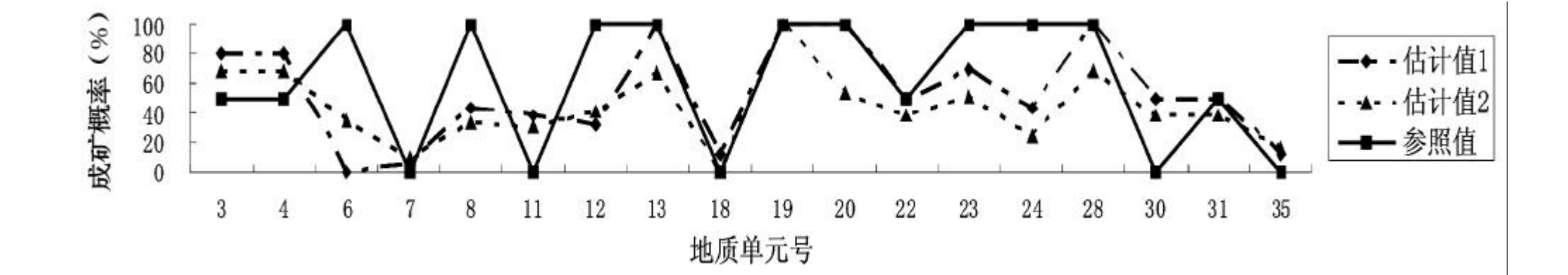


图 3 估计值与参照值对照图

Fig.3 Comparison diagram of estimate value and reference value

表 4 地质单元预测

Table 4 Prediction of geological cells

地质单元	关联度	地质单元	关联度
1	1.6417	25	1.9328
2	1.6417	26	1.9328
5	1.6417	27	1.1908
9	0.4509	29	1.4819
10	0.557	32	1.0847
14	1.0847	33	1.9328
15	1.0079	34	1.1908
16	0.9249	36	1.1908
17	1.9328	37	0.6338
21	0.6338		

(4) 预测。表 4 为基于最优变量的成矿单元预

测结果。

应用 K-Means Cluster 对模型单元关联度进行分析 ,如图 4 所示 ,得到聚类中心点为 Y =1.4819 ,即为临界值。若关联度大于该值 ,可认为有矿 ;小于该值 ,可认为无矿。预测结果为 :1 号、2 号、5 号、17 号、25 号、26 号、29 号和 33 号单元为成矿有利单元。矿产资源体的预测评价与周边的成矿背景场 (地质、物探和化探) 有很强的关联性 ,根据矿区大比例尺填图(1 2000)和勘查工程(1 500) ,预测有利单元中除 17 号和 26 号单元 ,其它地质单元均与研究区勘查工程资料相吻合。

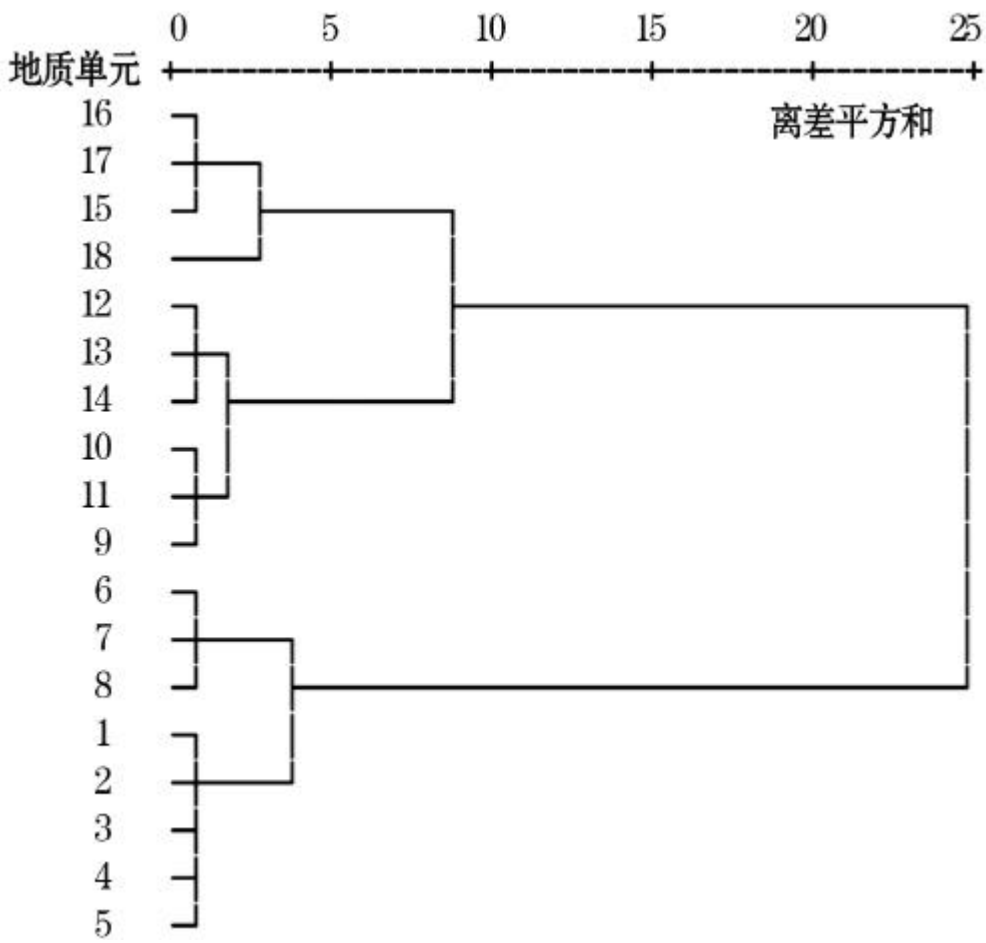


图 4 地质单元聚类分析图

Fig.4 Clustering diagram of geological cells

6 结 语

目前许多地学问题的研究中往往包含一些经验性的成分,所建立的数学模型具有似然性。为了有效地提取特征矿化信息进行成矿预测,必须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,把握数据分布的规律性。实验证明,粗糙集可以客观地筛选诸多变量,突出与成矿密切相关的信息,得到最佳变量组合,在此基础上构建特征函数,能够有效的降噪,找到最有利找矿标志或找矿标志的组合数值区间,为靶区预测提供更为准确的依据。

参考文献(References) :

[1] Pawlack Z. Rough Sets[J]. Communications of the ACM, 1995, 38

(11) 89-95.

- [2] Wang Xia, Tang Deshan. The evaluation method for regional water resources system based on roughness and fuzz set theory[J]. Water Resources Planning and Design, 2006, (1): 31-33. [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.]
- [3] Guo Jiayuan. Rough set-Based approach to rule generation and rule induction[J]. International Journal of General Systems, 2002, 31(6): 601-617.
- [4] Chu Yangjie, Wang Xianjia, Fang Debin, et al. An attribute reduction algorithm based on the relation matrix of rough set[J]. Journal of Wuhan University of Technology, 2006, 28(2): 81-83. [楚扬杰, 王先甲, 方德斌, 等. 基于粗糙集相关矩阵的属性约简算法[J]. 武汉理工大学学报, 2006, 28(2): 81-83.]
- [5] Pawalai Kraipeerapun, Chun Che Fung, Warwick Brown, et al. Uncertainty in mineral prospectivity prediction[J]. Neural Information Processing, 2006, 4233: 841-849.
- [6] Li Shuangcheng, Zheng Du. Applications of Artificial neural networks to geosciences: Review and prospect[J]. Advances in Earth Science, 2003, 18(1): 68-76. [李双成, 郑度. 人工神经网络模型在地学研究中的应用进展[J]. 地球科学进展, 2003, 18(1): 68-76.]
- [7] Xie Guiming, Fan Jizhang. The synthetic information model of prospecting gold deposits and prediction of prospecting gold targets in East HunChun District of Jilin Province[J]. Gold Science and Technology, 2000, 8(5): 20-27. [谢贵明, 范继璋. 吉林省珲春东部地区金矿综合信息找矿模型及找矿靶区预测[J]. 黄金科学技术, 2000, 8(5): 20-27.]
- [8] Huang Xiaonai, Chang Wensheng, Yang Jianming, et al. Mineral-ogenetic prediction of a certain uranium deposit[J]. Mining Research and Development, 2003, 23(6): 37-39. [黄晓乃, 畅文生, 杨建明, 等. 某铀矿床成矿预测[J]. 矿业研究与发展, 2003, 23(6): 37-39.]

Quantitative Analysis of Gold Mineral Resource Based on Rough Set

ZHU Yaqiong, YUAN Yanbin, ZHOU You, PENG Jingqian, ZHAN Yunjun

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

Abstract : Mineral information includes all kinds of relative metallogenic information. In order to extract comprehensive metallogenic prediction information, it's necessary to filter initial observation information to emphasize the factors which are most advantageous to metallogenic. Rough set can delete irrespective or unimportant attributes under the premises of no classification ability changing, without supplement information or prior knowledge. A new integrated prediction model based on Rough set theory is put forward in this research. The mineral information most advantageous to metallogenic from a great number of variables to achieve the optimization of variable structure and numerical interval is chosen. Based on the optimization combination, characteristic function is established for prediction. Combined with some conventional methods for deposit statistics, prediction, clustering means is applied to get the critical point for decision and quantitative characteristic analysis is applied to predict the mineral resource by calculating the relation degree of every geological cell. And eight geological cells are established as the cells advantageous to metallogenic. Results are basically in accord with practice, which shows availability of this method.

Key words : Mineral resource ; Rough set ; Quantitative prediction ; Characteristic analysis.