

37-42

秩特征分析方法在矿产资源预测中的应用

p612

陈永良 李学斌 许亚光

(长春地质学院·长春·130026)

提出了一种新的矿产资源靶区定位预测的统计方法—秩特征分析方法。该方法以地质变量之间秩相关分析为基础,根据地质变量集合中某一地质变量与其余地质变量之间总的秩相关程度来度量该变量的重要性大小,根据每个地质变量在统计单元(万能的资源靶区)上的取值情况计算单元成矿联系数,再根据单元成矿联系数相对大小评价优选矿产资源靶区。该统计方法可以同时使用定性、定量和半定量三种地质变量,减少了由于数据离散化而造成地质信息丢失,可以最大限度地利用各种类型地质变量所提供的有用信息。

关键词 秩特征分析, 地质信息, 资源靶区, 地质变量, 统计单元

矿产资源预测



技术·方法

特征分析方法^[1]及其改进后的统计方法(例如,分解加权特征分析^[2]、多维加权矩阵分析^[3])都要求输入定性数据,定量数据或半定量数据要离散化成定性数据才能使用。然而,在矿产资源预测中,常常存在一些特殊的定量或半定量地质变量,这些变量很难找到一个有确定地质意义的坎值来对它们作离散化处理,而人为的离散化处理会使原始数据中的大量信息丢失而不能得到充分的利用。同时,不同的离散化方案会得到不同的离散化数据,导致统计计算结果千差万别。

提出的秩特征分析方法既保留了特征分析方法的基本原理,又克服了特征分析方法要求使用定性数据的缺点,使定量数据、半定量数据和定性数据可以在模型中同时使用,使原始数据预处理工作大大减化^[4]。实践表明,秩特征分析方法用于矿产资源靶区预测是完全可行的。

1 秩特征分析方法的基本原理

秩特征分析与特征分析类似,都是传统类比法的一种定量化。其基本原理是通过研

究区已知矿床单元的研究,查明地质变量之间的内在联系,确定各个地质变量的成矿和找矿意义并赋给每一个变量特定的权系数,建立起某一类矿床单元的统计模型。然后,将该模型应用于预测单元,类比模型单元与预测单元在各种地质特征(地质变量)上的相似性,用它们之间的相似性程度来评价预测单元的成矿可能性。

设有 n 个矿床统计单元,有 m 个地质变量(其中包括定量地质变量、半定量地质变量和定性地质变量),统计单元的成矿联系数 $Y = (y_1, y_2, \dots, y_n)$ 可以定义为诸地质变量的线性组合。

$$Y = a_1 X_1 + a_2 X_2 + \dots + a_m X_m \quad (1)$$

式中, $X_1, X_2, \dots, X_m$ 表示 m 个地质变量, $a_1, a_2, \dots, a_m$ 是 m 个变量相应的权系数。 Y 是衡量统计单元之间相似性程度的一种度量,与特征分析方法相似,我们称之为统计单元成矿联系数。

秩特征分析方法要解决的首要问题是,根据模型统计单元的观测结果,以变量之间的秩相关分析^[5]为基础,确定变量在预测中的相对重要性程度,求得每个地质变量的权系数。

本文 1996 年 12 月收到,李翠华编辑。

1.1 原始数据预处理

定量地质变量和半定量地质变量的量纲和数量级往往千差万别,为了避免不同变量由于数量级不同而造成的预测结果的偏差(不作预处理会强调数量级较大变量的作用),在进行秩特征分析之前须进行数据预处理,以便统一数据的量纲和数量级。

设有同类矿床单元 n 个,变量 m 个。 m 个变量中包括定性变量、定量变量和半定量变量。变量 $X_j = (X_{1j}, X_{2j}, \dots, X_{nj})'$ 是定量或半定量地质变量,其预处理方法由(2)式给出。

$$X_{ij}' = \frac{X_{ij}}{\max_{1 \leq i \leq n} \{X_{ij}\}} \quad (i = 1, 2, \dots, n) \quad (2)$$

式中, X_{ij}' 是 X_j 预处理后的第 i 个分量。

定量、半定量地质变量预处理后,变量间的 τ -秩相关系数不变,变量各分量的取值在 $(0, 1]$ 区间上。

1.2 变量之间的秩相关性研究

1.2.1 τ -秩相关系数^[5]的定义

秩相关分析是研究两个变量秩之间相关性的一种统计分析方法。 τ -秩相关系数是根据两个变量秩之间“协调对”与“不协调对”的数目定义的。所谓协调对是指一个变量的两个观测值秩数的相对大小与另一个变量对应的两个观测值秩数的相对大小在变化方向上是一致的。非协调对是指两个变量相对应的观测值的相对大小在变化方向上是相反的。

设有两个变量 $X = (x_1, x_2, \dots, x_n)$ 和 $Y = (y_1, y_2, \dots, y_n)$, 如果在 i, j 两个样品上有 $x_i < x_j$ 且 $y_i < y_j$ 或 $x_i > x_j$ 且 $y_i > y_j$ 存在,那么,我们称 X 和 Y 之间存在一个协调对;否则,即称之为不协调对。把 X, Y 的各分量两两配对,统计 X 和 Y 之间协调对与不协调对的总数。令 P, Q 分别为 X 与 Y 之间协调对和不协调对的总数,那么, X 和 Y 之

间的 τ -秩相关系数定义为

$$\tau_{xy} = \frac{Z(P - Q)}{n(n - 1)} \quad (3)$$

τ -秩相关系数具有以下性质:

- ① $-1 \leq \tau_{xy} \leq 1$;
- ② $\tau_{xy} = 1$, 表示 x 与 y 的秩完全相同;
- ③ $\tau_{xy} = -1$, 表示 x 与 y 的秩完全相反;
- ④ $\tau_{xy} = 0$, 表示 x 与 y 的秩之间不存在任何相关性;
- ⑤ $\tau_{xy} > 0$, 表示 x 与 y 正相关;
- ⑥ $\tau_{xy} < 0$, 表示 x 与 y 反相关。

在实际计算中,可以采用以下的 τ -秩相关系数的一般形式:

设两个有序变量为 $X = (x_1, x_2, \dots, x_n)$ 和 $Y = (y_1, y_2, \dots, y_n)$ 。定义一种二元函数

$$(x_i, x_j) \rightarrow a_{ij}$$

$$(y_i, y_j) \rightarrow b_{ij}$$

要求

$$(1) \quad a_{ij} = -a_{ji}, \quad b_{ij} = -b_{ji};$$

$$(2) \quad a_{ii} = 0, \quad b_{ii} = 0.$$

那么, τ -秩相关系数的一般形式为

$$\tau_{xy} = \frac{\sum_{i,j} a_{ij} b_{ij}}{\sqrt{\sum_{i,j} a_{ij}^2 \sum_{i,j} b_{ij}^2}} \quad (4)$$

式中,两个二元函数 a_{ij} 和 b_{ij} 的取值方式作如下规定:

$$a_{ij} = \begin{cases} 1 & X_i < X_j \\ 0 & X_i = X_j \\ -1 & X_i > X_j \end{cases} \quad (i, j = 1, 2, \dots, n)$$

$$b_{ij} = \begin{cases} 1 & y_i < y_j \\ 0 & y_i = y_j \\ -1 & y_i > y_j \end{cases} \quad (i, j = 1, 2, \dots, n)$$

1.2.2 τ -秩相关矩阵的计算

设数据预处理后由 N 个模型单元数据构成的数据矩阵为

$$\tilde{X} = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1m} \\ X_{21} & X_{22} & \cdots & X_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ X_{N1} & X_{N2} & \cdots & X_{NM} \end{pmatrix} = (\tilde{X}_1, \tilde{X}_2, \cdots, \tilde{X}_m)$$

那么,任意两个变量 $\tilde{X}_j$ 与 $\tilde{X}_k$ ($j, k = 1, 2, \cdots, m$) 之间的 τ -秩相关系数可以根据公式(4)计算出来。计算出 m 个地质变量两两之间的 τ -秩相关系数,构成如下的 τ -秩相关矩阵

$$\tau = \begin{pmatrix} \tau_{11} & \tau_{12} & \cdots & \tau_{1m} \\ \tau_{21} & \tau_{22} & \cdots & \tau_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ \tau_{N1} & \tau_{N2} & \cdots & \tau_{NM} \end{pmatrix}$$

根据上述的 τ -秩相关矩阵,可以确定 m 个地质产量的权系数。

1.3 变量权的确定

仿照特征分析方法中变量权的求法,可用两种方法确定秩特征分析中地质变量的权系数。

(1) 矢量长度法

一般来说,一个变量与其它变量的相似性程度越高,其反映评价目标的信息也越多,那么,这个变量在评价中的重要性也越大。 τ -秩相关矩阵的每一行元素的平方和反映了各变量与其它变量的总相关程度,因此,可以用 τ -秩相关矩阵每行元素的平方和来定义各个变量的权系数。一般地,第 j 个变量的权系数为

$$a_j = \sqrt{\sum_{i=1}^m \tau_{ji}^2} \quad (j = 1, 2, \cdots, m) \quad (5)$$

(2) 主分量法

如果把 m 个地质变量的权系数看成是一个 m 维向量,那么,该向量代表了变量空间中某一特征方向,该特征方向反映了评价目标的总体特征,同时,该方向的特征性向量(由 m 个权系数构成的向量)应该与各变量最大程度地相似,即每个变量在该特征方向上的投影应尽可能地大。秩相关矩阵最大特

征值所对应的特征向量就具有上述性质,因此,可以作为权系数向量。

设 τ -秩相关矩阵(Γ)的最大特征值所对应的特征向量为

$$a = (a_1, a_2, \cdots, a_m)' \quad (6)$$

那么, $a_1, a_2, \cdots, a_m$ 即是为 m 个地质变量的权系数。

1.4 单元联系度的计算及预测单元评价

地质变量的权系数确定以后,根据方程(1)可以计算出所在矿床单元之间的成矿联系度。设 N 个模型单元的成矿联系度变化范围在 $y_1 \sim y_N$ 之间,检验模型单元的地质储量规模与单元联系度之间是否存在显著的相关关系。如果模型单元地质储量与成矿联系度之间存在显著的相关性,说明依据模型单元所建的统计模型是可靠的,可以对预测单元进行评价;否则,说明所建模型存在问题,应作进一步的研究。

2 秩特征分析在资源定位预测中的应用^①

牟-乳成矿带是胶东半岛重要的产金区之一。在该成矿带内,已发现的大、中、小型金矿床有十余处,金矿点及矿化点则星罗棋布,不胜枚举。为了查明牟-乳成矿带区域成矿规律及控矿因素,笔者等^②收集了大量有关牟-乳成矿带的研究资料,在该带内开展了地质、地球物理和地球化学等综合研究,根据十余处典型矿床特征的剖析,建立了牟-乳成矿带区域综合信息找矿模型。以矿构造为统计单元,在成矿带内圈定矿床统计

^① 国家黄金局地质科研项目《山东乳山县金青岭与江村洼矿区成矿预测》部分研究成果整理。

^② 参加该项研究的还有任洪茂和范继璋等。

单元 33 个,研究提取了包括地质、物探和化探等信息在内的地质变量 36 个,其中定量地质变量 3 个,半定量地质变量 7 个,定性地质变量 26 个。根据 11 个模型单元的观测数

据,建立了矿床靶区预测的秩特征分析模型,然后,根据相同的观测数据建立了靶区预测的独立权模型和特征分析模型。三种统计预测模型的计算结果见表 1 和表 2。

表 1 地质变量权系数统计表

类型	变 量		权 系 数				
	序号	名 称	独立权法	秩特征分析方法		特征分析方法	
				矢量长度法	主分量法	矢量长度法	主分量法
定 量	1	断 裂 数 目	0.7996	0.0547	0.4550	1.1427	0.4500
	2	交汇断裂有几个方向	1.000	0.0512	0.5396	0.8131	0.3736
	3	有几种指示元素异常	0.7981	0.0641	0.4786	0.4166	0.1403
半 定 量 变 量	1	单元内矿床规模	0.5859	0.0770	0.9747	2.0041	0.2615
	2	金元素异常强度	0.7996	0.0867	0.4696	1.1427	0.1718
	3	银元素异常强度	0.5802	0.0814	0.4696	2.6300	0.5499
	4	砷元素异常强度	0.5152	0.0848	0.4696	1.7675	0.3195
	5	金元素异常规模	0.5324	0.1114	0.1652	2.6473	0.0172
	6	银元素异常规模	0.5324	0.1158	0.0000	2.6473	0.2029
	7	砷元素异常规模	0.5324	0.1086	0.3797	2.6473	0.2926
定 性 变 量	1	存在金异常	0.0000	0.0000	0.4696	3.2676	0.2793
	2	存在砷异常	0.4615	0.1594	0.5425	3.1056	0.0000
	3	存在银异常	0.5359	0.1162	0.3560	2.8284	0.2222
	4	存在铜异常	0.7366	0.1296	0.3941	0.7714	0.2003
	5	存在蚀变带	0.0000	0.0000	0.4696	3.2677	0.2140
	6	控矿断裂为 SN 向	0.0000	0.0000	0.4696	3.2677	0.7228
	7	控矿断裂为 NNE 向	0.5108	0.1116	0.4696	0.5961	0.3968
	8	在 NE 向深断裂附近	0.5011	0.0843	0.4986	2.4546	0.1172
	9	在 SN 向深断裂附近	0.5108	0.0820	0.4613	0.6680	0.8075
	10	在 EW 向深断裂附近	0.0000	0.0000	0.4696	0.0000	0.3145
	11	在 NW 向深断裂附近	0.5288	0.1493	0.6328	0.3015	0.3316
	12	存在 NE 向断裂	0.7172	0.0752	0.4696	1.6812	0.4911
	13	存在 SN 向断裂	0.7172	0.1777	0.2919	2.9162	0.1611
	14	存在 EW 向断裂	0.0000	0.1777	0.4696	0.4360	0.1430
	15	存在 NW 向断裂	0.5892	0.0728	0.4696	1.2197	0.1094
	16	在岩体边界线性	0.5420	0.1122	0.8773	0.8907	0.3031
	17	在深部二导零值线上	0.7328	0.1016	0.1107	0.9664	0.1714
	18	存在元界地层	0.5420	0.1122	0.4696	0.8907	0.3542
	19	在岩体内接触带	0.5420	0.1122	0.4696	2.5713	0.3317
	20	在岩体外接触带	0.0000	0.0000	0.4696	0.0000	0.3145
	21	存在脉岩	0.0000	0.0000	0.4696	0.0000	0.3145
	22	处于 NE 与 NNE 向断裂交汇的锐角区	0.5806	0.0770	0.4345	2.2944	0.1969
	23	在航磁负场内	0.7366	0.1296	0.4696	0.7714	0.2037
	24	在航磁正场内	0.0000	0.0000	0.4696	0.0000	0.3145
	25	在航磁交变场内	0.7328	0.0837	0.6672	1.0718	0.3636
	26	在两种场分界线上	0.5253	0.0961	0.4696	1.6738	0.2460

注:特征分析模型使用的原始数据是秩特征分析模型所使用的原始数据用独立权法最优离散得到的。

由表 1 的统计结果可以看出,同一地质变量用不同的统计方法来计算变量权所得的结果差异很大,也就是说,同一地质变量在不

同的统计模型中的作用是不同的。这是否意味着不同的统计预测模型在评价同一地区会得出相反的结论呢?表 2 的统计结果回答了

上述问题。同表2中统计单元成矿联系度的统计结果可以看出,11个模型单元用三种统计模型计算成矿联系度,模型单元成矿联系度的排列顺序是大体相同的,成矿联系度最大的是金青顶,最小的是初家沟;22个预测单元的成矿联系度统计结果也表明,秩特征

分析方法与特征分析方法及独立权法的评价结果基本相似。取三种模型评价结果的部分,即可以圈定出牟-乳成矿带内的矿床远景靶区,这样的远景靶区共有6个,单元序号分别是13、15、19、31、32、33。

验证三种统计方法评价结果的可靠性,

表2 矿床统计单元成矿联系度统计表

统计单元		成 矿 联 系 度					评 价
序号	名 称	独立权法	秩特征分析方法		特征分析方法		
			矢量长度法	主分量法	矢量长度法	主分量法	
1	金青顶	13.0712	2.030	10.179	46.345	6.042	M
2	邓格庄	11.8068	1.703	9.389	46.585	5.884	M
3	西直格庄	7.1127	1.544	8.581	35.656	3.865	M
4	铜锡山	4.8371	1.436	6.903	24.762	3.082	M
5	富禄地	8.7870	1.748	8.077	40.446	4.444	M
6	金牛山	9.0371	1.680	8.766	41.705	5.299	M
7	初家沟	3.7600	0.932	6.004	19.849	2.512	M
8	唐家沟	7.6532	1.751	9.766	34.437	5.329	M
9	哈狗山	7.6924	1.517	8.105	41.389	4.257	M
10	岔 河	6.4119	1.503	7.684	36.984	3.551	M
11	山上里	6.5259	1.402	7.210	37.090	3.799	M
12	三 甲	4.1558	0.905	6.009	23.397	3.309	
13	胡家口	9.2072	1.581	9.171	32.017	5.736	T
14	胡八庄	5.4139	1.035	5.435	22.770	3.133	
15	金青岭	7.7345	1.776	9.221	31.419	4.135	T
16	战家芥	5.4234	1.162	6.354	20.307	2.980	
17	洼九布	2.7844	0.552	3.546	15.378	2.172	
18	仇家洼	5.3394	1.001	5.394	23.047	2.910	
19	清口涧	6.9979	1.579	8.210	30.648	3.934	T
20	赵 家	3.8432	0.912	6.225	18.314	3.077	
21	江村洼	3.9954	0.789	5.422	17.092	2.929	
22	岚 子	3.2974	0.754	4.548	18.655	1.915	
23	南江村	5.1698	0.982	4.899	18.422	2.331	
24	大帝庙	3.1422	0.485	4.125	15.534	2.350	
25	围子山	2.5849	0.416	3.142	12.779	2.073	
26	吉林村	2.5812	0.370	3.810	13.079	2.547	
27	上 册	1.8359	0.338	3.139	12.633	1.749	
28	文史家	3.0783	0.535	3.644	15.883	2.294	
29	丛 家	3.6423	0.712	4.502	17.476	2.509	
30	大王口	2.1642	0.629	4.421	19.331	2.133	
31	东立芥南	6.8967	1.409	7.969	29.465	4.150	T
32	福 山	4.2072	0.989	6.595	21.323	3.271	T
33	官 庄	6.9979	1.553	6.871	24.112	2.998	T

注:表中M代表模型;T代表靶区。

可以用德尔菲法的评价结果与之比较。但因客观条件所限,本文仅采用简单而实用的“模型回代”法来验证评价结果(表3)。

从表3的相关系数统计结果可以看出,用三种统计方法计算出的成矿联系度与矿床工业品位、工业储量及成矿潜力之间存在密

表 3 模型单元回代结果检验表

单元号	平均品位	工业储量	成矿潜力	与成矿联系度之间的相关系数				
				独立权法	秩特征分析方法		特征分析方法	
					矢量长度法	主分量法	矢量长度法	主分量法
1	18.56	29.00	538.20					
2	8.73	29.24	225.27					
3	8.33	8.40	70.00	0.680	0.698	0.550	0.424	0.511
4	10.19	2.98	30.40	0.790	0.498	0.604	0.520	0.648
5	10.00	2.50	25.00	0.780	0.599	0.615	0.501	0.620
6	7.50	>3.00	22.50					
7	4.50	4.69	21.10					
8	6.00	2.72	16.32					
9	6.00	1.89	11.32					
10	7.00	1.47	10.28					
11	4.00	0.624	2.50					

注:表中的三行相关系数分别指平均品位、工业储量、成矿潜力(品位×储量)与成矿联系度之间的相关系数。

切的相关性,成矿联系度在一定程度上反映了矿床的成矿综合潜力。可见,秩特征分析方法用于资源靶区预测是完全可行的。

秩特征分析方法与一般特征分析方法相比其主要优点在于,它可以同时使用定量、半定量和定性三种类型的地质变量来进行矿产资源靶区预测。因此,在矿产资源靶区预测中,秩特征分析方法较一般特征分析方法更有适应性,更具有应用前景。

参考文献

- 1 Botbol J. M., An application of characteristic analysis to

mineral exploration, Proc. gth. Int., 1971

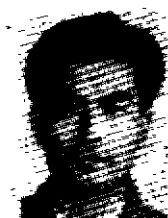
- 2 Pan G C, and Harris D. P., Decomposed and weighted characteristic analysis in the quantitative evaluation of mineral resources with a case study on the pegmatic Nb - Ta deposits in China, Math. Geol., 1989
- 3 刘安洲,范兴业,刘少华,成矿信息论与大比例尺矿产靶区定量预测,长春:大学生出版社,1991
- 4 陈永良,杜德文,刘安洲,独立权法及其在矿产预测中的应用,长春地质学院学报,第三十届国际地质大会论文集,1996
- 5 Kendall M. G., Rank correlation methods, Charles Griffin & Company LTD., 1970

APPLICATION OF RANKED CHARACTERISTIC ANALYSIS TO MINERAL RESOURCES PREDICTION

Chen Yongliang, Li Xuebin, Xu Yaguang

A new statistical method called ranked characteristic analysis based on the rank analysis among geological variables for target prediction of mineral resources was presented. The significance of one variable was determined by the ranked correlativity between the variable and the others. Mineral resources targets can be selected by the metallogenic favorabilities of statistical cells. In this model, qualitative, quantitative and semi-quantitative geological variables can be used simultaneously, and useful information of geological variables can be used completely.

Key words ranked characteristic analysis, geological information, resource target, prediction



第一作者简介:

陈永良 男,1965年生。1989年毕业于长春地质学院地质系地质专业,1992年在长春地质学院基础科学系获硕士学位。现为长春地质学院综合信息矿产预测研究所讲师,主要从事数学地质方面的科研和教学工作。

通讯地址:吉林省长春市长春科技大学(原长春地院) 综合信息矿产预测研究所 邮政编码:130026