

如事物的因果规律性很明显,样本数大于10也能保证一定的可靠性。在选用统计数据时,注意剔除明显的非正常因素是必要的(如个别钻孔孔内情况特殊、复杂,钻进事故很多,这种钻孔的统计资料就不宜参与统计计算)。其二,这种方法所得的结果,在变量的平均数周围的一定区间内,其预测的准确程度高,不适当的外延是不许可的。比如,用勘探矿区平均岩石可钻性0.6~2.5(大致相当于4~7级岩石),孔深平均250~550米范围内的生产统计数据所建立的回归方程,不一定能适用于普查浅钻的生产。其三,必须严格注意统计资料的来源具有一致性。如小口径钻进的生产统计资料,就不能与大口径钻进的生产统计资料混在一起作数学处理。其四,统计数据必须真实,否则,假帐真算,将会导致错误的结论。

本文承湖南、湖北、广西地矿局提供资料,谨致谢意。

特征分析法简介及其初步应用

西北冶金地质勘探公司地质研究所 雷祖志 姜明辉

1981年5月,联合国教科文组织在北京举办的矿产资源评价培训班上,麦克卡门(R. B. McCammon)等人介绍了特征分析法。它是通过建立反映某一类型矿床的特征模型,来达到预测该类型矿床的目的。

我们对该方法的基本思想、原理及其应用条件等进行了研究,并应用于陕西小秦岭地区钼矿的统计预测,取得了较好的效果。现简介如下。

基本思想

特征分析的基本思想是从已知某一类型的矿床中,通过一定的数学方法找到该类型矿床的“特征”。这种“特征”是该类型矿床的共性表现,它反映该类型矿床特征的地质因素间的特定组合。这种特定组合表现为若干个特征变量的最佳加权线性组合,即特征模型。藉此模型可在未知区寻找具有这种“特征”的同类型矿床。

加权线性组合怎样才算最佳呢?这必须具备两点:1)加权线性组合中必须包括该类型矿床的确定性因素,或者线性组合必须寓于该类型矿床的确定性因素之中。例如斑岩体就是斑岩型矿床的确定性因素;对层控多金属矿床来说,某一特定的地层层位就是其确定性因素;2)尽可能满足如下函数关系:

$$y = a \sum_{i=1}^m x_i$$

(1) 式中 x_i 为特征变量, a 为 x_i 的权系数, y 为关联度。

用图形表示如图1。

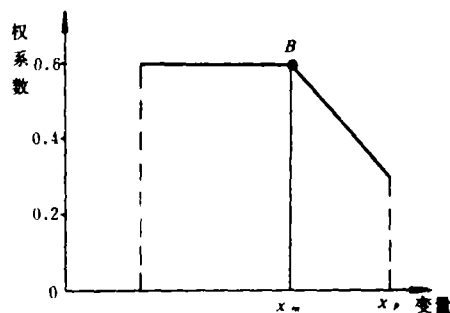


图 1

那么特征模型应由 B 点以左的 m 个变量组成,这 m 个变量就称为特征变量。

为什么必须具备以上两点呢?显而易见,如果1)不满足,例如要建立斑岩型矿床的特征模型,如果不是在与斑岩有成因联系的矿床中建立(而是在岩浆矿床中),这样的特征模型显然不具备斑岩矿床的特征;同样,如果层控矿床没有特定的地层层位,也就不能称其为层控矿床。所以,特征模型必须包括确定性因素或寓于其中。这是评价特征模型好坏的最重要标准。至于2),特征变量的权系数最好相等(即诸变量在逻辑矩阵中的构形完全相同,也就是最特征),但往往有所差异,这是允许的。但随着差异增大,特征模

型的特征性将逐步减弱。

特征分析还有一个重要的特点,就是它在寻找隐伏矿床过程中揭示了“元素相对浓度比绝对浓度意义更大”^[6]。由于出露地表的易发现的矿床愈来愈少,目前趋向于应用物化探方法来寻找盲矿床,而特征分析在处理物化探数据方面有其独到之处。

麦克卡门等人对美国内华达州伊利(Ely)附近的一个斑岩铜矿,应用特征分析建立了如下模型: $y = 0.49As + 0.47Sb + 0.46Ag + 0.43Zn + 0.38Fe$ 。该模型是在斑岩铜矿床上根据37个元素(化探原生晕)的绝对浓度变成相对浓度后建立的。虽然该模型中不包括斑岩这个确定性因素,但是它寓于其中;各特征变量的权系数虽略有差异,但仍为一个比较好的特征模型。因此,他们应用该模型在罗坎扬(Rowe Canyon)地区找到了一个隐伏斑岩铜矿床。

方法及原理

特征分析定义为特征变量的最佳加权线性组合,权系数是由特定的控制单元内特征变量之间的相互关系确定的。模型的建立分以下几步:1)选择控制单元;2)原始数据的变换;3)模型特征定量化;4)最佳特征变量集的确定。

1. 选择控制单元 首先选择一定大小的网格,将整个研究区划分为等面积、同形状的若干单元(或以地质体等为单元)。这些单元应满足以下条件:

1) 研究程度较高,具代表性,能最大限度反映某类型矿床的特征;

2) 控制单元数要尽可能大于变量数,以使建立的特征模型具统计意义;

3) 要考虑到预测单元的实际情况,使建立的模型能在预测区应用。

2. 原始数据的变换 特征分析是以各控制单元中各变量间的匹配关系为计算基础的。它通过对比控制单元中各变量匹配数与不匹配数来研究变量对之间的相关关系,进而揭示某类型矿床的特征。为此,必须将控制单元各变量的原始数据转换成逻辑变量。

所谓逻辑变量,就是表示地质因素逻辑上的对立的各状态,如某地质因素的存在或缺失,某异常的高与低等对立概念。逻辑变量可以是二态的,即“0、1”;也可以是三态的,即“+1、0、-1”。当逻辑变量为二态时,0、1分别表示其变量在逻辑上的两种对立状态;当逻辑变量为三态时,0则表示中间状态。显然,三态变量比二态变量更能恰当地反映各种地质现象。

要把原始数据化为逻辑变量,首先应根据变量的性质将其区分为“定态变量”和“不定态变量”。所谓定态变量是指该变量在全研究区内各单元中的取值是确定的,如化探分散流Pb元素的含量,各单元都有确定的取值;所谓不定态变量是指该变量在全研究区内只有部分单元取得数值,而另一部分单元则不能取得确定数据,例如,以斑岩体的有无作为变量,显然,有斑岩体出露的单元可以取得数据(如+1),而未见斑岩体出露的单元却不能取得确定的值(如-1),因为这些单元可能有隐伏斑岩体存在。

一般而言,定态变量多为连续型变量,不定态变量多为离散型变量;就数据性质而言,前者多为定量数据,后者多为定性数据。

区分变量状态以后,对定态变量要将绝对“浓度”化为相对“浓度”。相对浓度值可以是连续型数据,也可以是离散型数据,但最后都必须变换成逻辑变量。变换方法较多,至于采用哪种方法,要紧密结合地质概念模型,根据具体情况选择合适的变换方法。通常采用的方法有:衬度法、邻近单元比较法^[1]、求二阶导数法^[5]等。

如果将定态变量与不定态变量混合计算,必须把两者统一到相同的意义上。这一点十分重要。例如定态变量中,正异常为+1,无异常为0,负异常为-1,这与不定态变量中的+1表示某变量的存在,-1表示不存在,0表示性质不明的含义完全不相同。显然,以这样含义不同的匹配关系来计算,将会导致错误的结论。

那么,如何统一呢?假定已将不定态变量转换成三态逻辑变量,+1表示某变量存在,0表示性质不明,-1表示缺失。对于定态变量,可选用某一恰当的方法,将其划分成若干个不同的

区间,视各区间为不同的变量,某变量在单元中出现为+1,缺失为-1。对动态变量来说,表示情况不明的0不存在。

到此,得逻辑变量矩阵A。

$$A = a_{ij} (i = 1, 2, \dots, n; j = 1, 2, \dots, p) \quad (2)$$

式中, n 为单元数, p 为变量数。

3. 模型特征定量化 对某一类型矿床而言,从许多地质因素中寻找反映某特征的地质因素,并对这些特征因素以某种数学方法加以定量描述,称为模型特征定量化。模型特征定量化通常采用以下三种方法: 1) 平方和法; 2) 乘积矩阵主分量法; 3) 概率矩阵主分量法。

下面仅讨论乘积矩阵主分量法。

假设已将原始数据转换成逻辑变量矩阵,并用A表示。

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{np} \end{bmatrix} \quad (3)$$

用A左乘其转置矩阵A'得乘积矩阵 $B_{p \times p}$

$$B = A' \cdot A \quad (4)$$

B阵中对角线项是每个变量自身正一正匹配数与负一负匹配数之和;非对角线项为两变量匹配数与不匹配数之差。 i, j 两变量匹配数与不匹配数之差越大,说明 i, j 两变量在所有控制单元中关系越密切。

若把B阵的各列或各行视为一向量,那么每一个向量则表示每一个变量与所有其他变量之间总的匹配情况。特征分析就是为了对比有意义的各变量的匹配关系,从中找出若干个相关程度最好的变量组成特征变量集。由于这一变量集在所有控制单元中共同出现,因此它也就反映了所有控制单元的共同特征。

乘积矩阵主分量法是在B阵的基础上进行如下运算:

1) 求出B阵在实数域内所有的特征值 λ_i ($i = 1, 2, \dots, p$)。

2) 求出 $\lambda_{\max}$ 所对应的全部特征向量的基础解系 $X_{p \times 1}$ 。

特征向量 $X_{p \times 1}$ 的各元素就可作为对相应变

量的定量描述。这是否可行呢?我们分析一下特征值及特征向量的几何意义:

$p \times p$ 矩阵,可看成是 p 维空间中点的坐标的一种表示。例如, 2×2 矩阵可看成是平面上的两个点。如矩阵 B_1 :

$$B_1 = \begin{bmatrix} 4 & 8 \\ 8 & 4 \end{bmatrix}$$

可看作是坐标平面上的两个点: $x_1(4, 8)$, $x_2(8, 4)$ 。这就定义了以坐标原点为起点的两个向量。可设想这两个点位于以坐标原点为中心的椭圆上。那么,特征值就代表此椭圆的长半轴和短半轴。其特征值为 $\lambda_1 = 12$, $\lambda_2 = -4$ [$|\lambda_1| / (|\lambda_1| + |\lambda_2|) \times 100\% = 75\%$]。对应于 λ_1 和 λ_2 的特征向量分别为 $(1, 1)'$, $(-1, 1)'$ 。第一特征值代表主轴,自中心至边缘的长度为12单位,第二特征值代表副轴,位于坐标系的负区,长度为4单位(图2)。

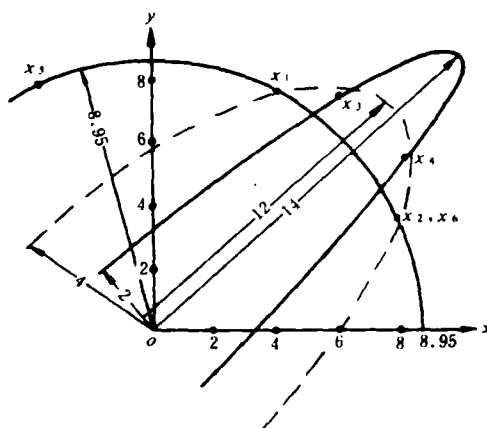


图 2

当两点数为靠近时, λ_1 与 λ_2 的值就发生变化,长轴变长,短轴变短。如矩阵 B_2 :

$$B_2 = \begin{bmatrix} 6 & 8 \\ 8 & 6 \end{bmatrix}$$

亦可看成两个点: $x_3(6, 8)$, $x_4(8, 6)$,其特征值 $\lambda_1 = 14$, $\lambda_2 = -2$ [$|\lambda_1| / (|\lambda_1| + |\lambda_2|) \times 100\% = 88\%$],以图2中实线椭圆表示。

当两点重合,则两向量重合,包围的椭圆就成为一条直线,长轴达到最大值,短轴为0。反之,如果两向量正交,则椭圆变成圆,向量的长即圆的半径,亦即两特征值的绝对值。如矩阵B:

$$B_3 = \begin{bmatrix} -4 & 8 \\ 8 & 4 \end{bmatrix}$$

可看作点 $x_5(-4, 8)$ 和 $x_6(8, 4)$, 特征值 $\lambda_1 = +8.95$, $\lambda_2 = -8.95$ [$|\lambda_1| / (|\lambda_1| + |\lambda_2|) \times 100\% = 50\%$], 此关系如图 2 所示。

以上情况可扩展到 $p > 2$ 的情况, 即 $p \times p$ 矩阵的元素可视为 p 维椭球体上的点, 矩阵的特征向量产生椭球体的主轴, 而特征值代表这些轴的半长。主分量法就是要寻找这些轴并度量其长度。就乘积矩阵 B 而言, 因为它是一个 $p \times p$ 阶的匹配矩阵, 所以可计算 p 个特征值和 p 个特征向量。由于匹配矩阵总是对称的, 因此 p 个特征向量互相垂直。我们可把每一特征向量都看作是一个新的变量, 把特征向量各元素看作是相对应变量在新变量上的“载荷”。这一点与因子分析的思想是相同的: 不同的是, 特征分析只选取最大特征值所对应的特征向量来建立特征模型。所以, 必须保证最大特征值在特征值总和中占较大的百分比, 该比值越大说明 p 维空间中各变量点之间的相关程度越好。从以上三个简单矩阵可以看出, 变量之间的关系越密切, 则最大特征值所占总和的百分比越大, 当两变量完全无关时, 比值最小。因此, 可以根据该百分比的大小来衡量特征模型的好坏。同时, 该比值越大也说明提取包含在诸特征变量中的找矿信息越多, 但前提是最大特征值必须是“成矿因子”。

4. 最佳特征变量集的确定 运用主分量法对各变量赋权以后, 将各变量权系数由大到小排列, 作出变量与其权系数的关系图, 并结合地质概念模型, 选择截取点 B , 以 B 点所指示的横坐标以左的各变量作为特征变量 (见图 1)。然后, 以所选取的各特征变量在所有控制单元中的匹配关系重新对其赋权。这一过程是很重要的, 因为

以前对变量所赋的权系数是根据所有参与计算的变量之间的匹配关系来确定的, 它反映所有变量之间的相互关系。但是, 特征变量之间的匹配关系还应由它们之间的匹配关系来确定。这样, 就可以得到最佳的特征模型。

应用实例

在陕西小秦岭地区的钼矿统计预测中, 应用特征分析法, 进行了斑岩钼矿床的统计预测。现将主要结果概述如下:

在已有条件下, 选择 7 个已知斑岩钼矿床所在单元作为控制单元。根据地质认识以及多种统计方法选择与成矿有关的 10 个地质因素:

- 1) D_1 —单元中心与 I 级深成断裂的最短距离 (km);
- 2) D_p —单元中心与北东及北西向深成断裂交汇点的最短距离 (km);
- 3) Sc —单元落在卫片偏心环影像内的面积 (km^2);
- 4) Bl —单元内燕山期中酸性大岩基和小岩体是否同时出现;
- 5) Pc —单元内燕山晚期花岗斑岩体的岩石化学成分综合变量;
- 6) $Al-A$ —围岩蚀变;
- 7) Wg —重砂异常;
- 8) Ms —航磁异常;
- 9) Mo —化探分散流钼元素在单元中的平均含量 (ppm);
- 10) Cp —化探分散流综合变量。

以上 10 个变量中, Bl 、 Pc 、 $Al-A$ 为不定态变量, 其余为定态变量。对定态变量化绝对“浓度”为相对“浓度”, 再选择部分无矿单元与已知含矿单元用 t 统计量确定各变量转换为逻辑变量的临界值 (表 1)。

表 1

变量	D_1	D_p	Sc	Bl	Pc	$Al-A$	Wg	Ms	Mo	Cp
临界值	<4.2 为 +1, 否则为 -1	<7.55 为 +1, 否则为 -1	落入环内为 +1, 否则为 -1	出现为 +1, 否则为 0	≥ 5.58 为 +1, 否则为 0	≥ 4 为 +1, 否则为 0	出现为 +1, 否则为 -1	出现为 +1, 否则为 -1	≥ 3.85 为 +1, 否则为 -1	<0.92 为 +1, 否则为 -1

根据表1将10个变量在控制单元中的数据转换为三态变量得逻辑矩阵(表2)。并得乘积矩

阵(表3)。以乘积矩阵最大特征值 λ_{max} 所对应的特征向量 X 表示各变量的权系数(表4)。

表2

单元 \ 变量	D_1	Ms	Dp	Sc	Bl	Pc	$Al-A$	Wg	Mo	Cp
1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1	1
4	1	-1	1	1	1	1	1	1	1	1
5	1	-1	1	1	0	1	1	1	1	1
6	1	-1	1	1	0	1	0	1	1	1
7	1	1	1	1	0	1	1	1	1	1

表3

匹配数 变量 \ 变量	D_1	Ms	Dp	Sc	Bl	Pc	$Al-A$	Wg	Mo	Cp
D_1	7	1	7	7	4	7	6	7	7	7
Ms	1	7	1	1	2	1	2	1	1	1
Dp	7	1	7	7	4	7	6	7	7	7
Sc	7	1	7	7	4	7	6	7	7	7
Bl	4	2	4	4	4	4	4	4	4	4
Pc	7	1	7	7	4	7	6	7	7	7
$Al-A$	6	2	6	6	4	6	6	6	6	6
Wg	7	1	7	7	4	7	6	7	7	7
Mo	7	1	7	7	4	7	6	7	7	7
Cp	7	1	7	7	4	7	6	7	7	7

表4

变量	D_1	Dp	Sc	Pc	Wg	Mo	Cp	$Al-A$	Bl	Ms
X	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.31	0.21	0.07

因 $(|\lambda_{max}| / \sum_{i=1}^m |\lambda_i|) \times 100\% = 86.21\%$,故可采用 X 各元素作图(图3)。

以各特征变量在控制单元中的匹配矩阵为基础,采用乘积矩阵主分量法最后确定各特征变量的权系数,建立斑岩钼矿床的特征模型为:

$$y = 0.359 \times (Mo + Pc + Cp + D_1 + Wg + Sc + Dp) + 0.313 \times Al-A$$

模型中包括确定性变量 Pc ,并有 $(|\lambda_{max}| / \sum_{i=1}^m |\lambda_i|) \times 100\% = 98.6\%$,故可采用此模型在未知区进行斑岩钼矿床的预测工作。

首先将预测单元以与控制单元完全相同的原

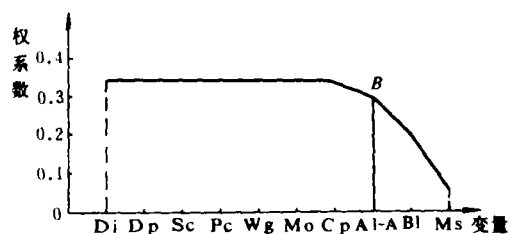


图3

则,把诸特征变量的原始数据化为逻辑变量,然后统计预测单元与模型的匹配数。当单元中特征变量全部显示+1时与模型完全匹配,显示+1少于8的单元,其匹配程度相应降低,与模型的相似程度也相应减少。统计结果见图4。

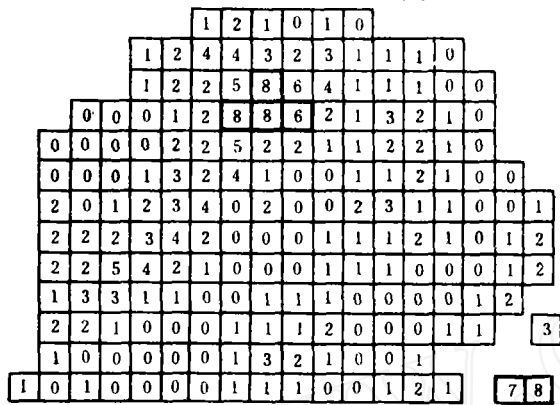


图4 小秦岭地区斑岩型钼矿床预测示意图
方格中数字为与特征模型匹配数,
粗线方格为已知矿床所在单元。

从图4可见,与模型匹配数大于4的单元共有17个,包括全部已知含矿单元。故与模型匹配数大于4的单元为斑岩型钼矿床成矿有利单元。

(上接第23页)

的黄铁矿具有一致的硫同位素组成和分馏特征,以及Au、Ag、Ni、Co含量相对较低,Co/Ni

根据特征模型,还计算了部分单元(包括7个已知含矿单元及30个无矿单元)与模型的关联度 ν ,以 ν 值计算了与已知矿床规模大小的拟合度,结果高达93.54%;同时,我们把一个已知含矿单元留作预测验证,结果以较高的相似度(匹配数为6)预测为含矿单元。这就说明该模型具有一定的可靠性。从图4可见,预测单元及已知含矿单元明显地呈北东向展布,并略有北西向趋势,这和构造岩浆活动带是本区钼矿成矿带的重要控制因素这种地质认识相一致。

特征分析法是一种较新的统计分析方法,在研究该方法的过程中得到赵鹏大、黄世乾、胡旺亮、周宏坤等同志的热情指导,在此表示衷心感谢。

参考文献

- [1] 赵鹏大等,矿床统计预测,1981(未刊稿)
- [2] 中国科学院地质研究所,数学地质引论,1976,地质出版社
- [3] 於崇文等,数学地质的方法与应用,1978,冶金工业出版社
- [4] 谢帮杰,线性代数,1978,人民教育出版社
- [5] 李纯杰等,特征分析,1981(未刊稿)
- [6] R. B. 麦克卡门,冶金地质动态,1982,第8期

不同类型地质体黄铁矿Au、Ag、Ni、Co含量						表3	
地质体类型	样品数	平均含量 (ppm)				Ag/Au	Co/Ni
		Au	Ag	Ni	Co		
矿体内细粒浸染—星散状硫化物	7	117.74	695.36	91.57	71.29	5.90	0.78
矿体内粗晶团块—脉状硫化物	8	0.60	56.36	13.25	33.63	93.93	2.54
围岩(陈蔡群)	5	8.50	65.60	383.20	205.16	7.72	0.54
火山、次火山岩中多金属硫化物脉	3	0.33	14.00	4.70	20.33	42.42	4.33

如前所述,本区金银矿体中并存两种不同产出形态,不同成因的未经同化均一的硫。细粒浸染—星散状硫化物硫来自围岩(陈蔡群变质岩),系加里东旋回变质热液流体活化围岩中的硫、金、银等元素,向同期构造变动所形成的低压低位带——控(容)矿断裂构造迁移、沉淀的产物;粗

>2.5的相同特点(见表3),推断它们属于同一热源,系火山期后热液分异的含矿流体。

晶团块—脉状硫化物硫与燕山旋回火山期后热液分异流体有关。前者与金银矿化关系密切,是金银的主要载体,后者含少量银、微量金。据此认为,本区金银矿床的主体成矿时代为加里东期,燕山期又在已成矿体的背景上叠加了少量银、微量金矿化。