第25卷 第5期 地质与勘探 Vol.25,No.5
1989年5月 GEOLOGY AND PROSPECTING May, 1989

一种用于含矿性评价的双重分析方法*

杨 永 华

（长春地质学院）

本文讨论一种改进的特征分析方法。模型中采用一组非矿样品作对照，从两个相反的角度表现矿的特征。讨论了用矿的非矿关联度进行样品综合标度的形式和建立对象权理论分布的意义，研究了通过样品对比实现判别归类和给出含矿性估计的近似概率方法。

关键词：矿产资源；评价；概率

方法原理

在矿产资源定量评价中，为了解决样品（地质单元）含矿性估计问题，常用特征分析方法。它是一种使用定性数据、从变量之间匹配关系出发、研究样品在一维空间中标度性质的方法。它通过对一组含矿样品的综合，找出典型特征方向作为矿化特征来建立矿床模型。这个方法对预测样品的含矿性估计是通过样品关联度与模型进行对比来实现的。但是一般特征分析的此类对比是单向的，即它仅根据预测对象与矿床的相似程度来决定样品含矿性，因而具有一定的片面性，因为，与矿床地质条件相似的地质体并不一定都成矿。根据这种考虑，本文提出一种改进的方法，它对模型样品同时进行矿和非矿的双重标度，并通过它们之间关系的研究，从正、反两个方面来表现模型的特征。与一般特征分析的区别是这个方法对照使用了一组非矿模型样品，它将模型的非矿特征也作为对比的依据。对预测样品的含矿性估计是通过一个综合标度值——对象权与模型进行对比来实现的判别归类。分类的原则是：一个含矿可能性大的预测样品，其标度性质与矿相似而与非矿不相似。

在研究此法前，首先给出三个基本假设：①变量二态赋值遵循有利原则，即为变量的状态对矿有利时赋值为1，否则赋值为0；②在样品上有利变量相配合是矿化信息的反映；③矿化信息强度与含矿有利性成正比。

数学模型

设有矿和非矿两类共$s+r=n$个样品，它们符合作为两个母体随机样本的要求。在全部n个样品上对m个地质变量进行观测，用1表示“存在”，用0表示“不存在”这两种状态，得到两个原始赋值矩阵$\boldsymbol{X}$和$\boldsymbol{Y}$：

$$\boldsymbol{X}=\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ x_{s1} & x_{s2} & \cdots & x_{sm} \end{pmatrix}$$

* 国家自然科学基金资助课题。

$$Y=\begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1m} \\ y_{21} & y_{22} & \cdots & y_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ y_{r1} & y_{r2} & \cdots & y_{rm} \end{pmatrix},$$

矩阵中的行代表样品，列代表变量。将X、Y按列分块，记为：

$$X=(x_1,\ x_2,\ \cdots,\ x_m)$$

$$Y=(y_1,\ y_2,\ \cdots,\ y_m),$$

其中，x_j、y_j是第j个地质变量的向量表示。

取矩阵Y的补矩阵$\widetilde{Y}$，

$$\widetilde{Y}=(\widetilde{y}_1,\ \widetilde{y}_2,\ \cdots,\ \widetilde{y}_m),$$

$\widetilde{y}_j$是定义为$\widetilde{y}_j=1-y_j$的补变量，其中1是取值全为1的r维列向量。显然，$\widetilde{y}_j$亦是由二态值0、1构成，它可理解为与x_j定义相反的一个新变量。

变量取补的目的是将两类样品中变量的作用统一。由于X和Y中变量定义相同而状态值含义相反，前者用1表示有利于矿而后者用0表示有利于矿，根据前述变量赋值的有利性假设，为利用非矿样品中“0”所提供的信息，因而对非矿赋值矩阵进行了取补处理。这样，对第j个变量重要性的衡量可通过它在X和$\widetilde{Y}$中的表现来研究，即将存在此变量时指示矿的作用和不存在此变量时指示非矿的作用综合考虑。

现在考虑用X和$\widetilde{Y}$两套数据分别建立模型：

$$u=X\alpha=\alpha_1x_1+\alpha_2x_2+\cdots+\alpha_mx_m$$

$$v=\widetilde{Y}\beta=\beta_1\widetilde{y}_1+\beta_2\widetilde{y}_2+\cdots+\beta_m\widetilde{y}_m,$$

u、v分别是矿和非矿样品的得分向量。构造模型的基本思想是：在变量空间矿样品点群中寻找一个方向α，使样品点在此方向上的得分u_j都较大；同时在非矿样品点群中也寻找一个方向β，使非矿样品在此方向上的得分v_j也都大。

对前一个模型，问题可表述为：寻找一组数$(\alpha_1,\ \alpha_2\cdots,\ \alpha_m)=\alpha$，使线性组合$u=\alpha_1x_1+\alpha_2x_2+\cdots+\alpha_mx_m$在$\|\alpha\|$一定时长度最大。由于$\|u\|^2=u'u$与$\|\alpha\|$有相似的度量性质，因此这个准则相当于求$\alpha$使$u'u=\alpha'X'X\alpha$在$\|\alpha\|=1$条件下取最大值。由拉格朗日乘数法，求：

$$Q=\alpha'X'X\alpha+\lambda(\alpha'\alpha-1)$$

的一阶偏导数，令

$$\frac{\partial Q}{\partial\alpha}=0,$$

便有

$$2X'X\alpha-2\lambda\alpha=0,$$

即

$$X'X\alpha=\lambda\alpha$$

于是问题归结为求矩阵$X'X$的特征值和特征向量，当λ取为最大特征值时，相应的特征向量α使$u'u$取最大值。

类似地，后一个模型也可通过求$\widetilde{Y}'\widetilde{Y}$的最大特征值找到一个特征向量$\beta$，使$v'v$取最大值。

矩阵$X'X$一般称为变量匹配矩阵，它反映变量所代表的各种地质条件在矿上的有利配合关系。用这种匹配表现变量对样品矿化的作用具有度量上的一致性：出现有利变量匹配越多的样品越有利于成矿。而矩阵$\widetilde{Y}'\widetilde{Y}$表示不利变量在非矿上的配合，它从相反方面表现了各变量对矿的上述度量性质：不利变量配合越多的样品越不利于成矿。

在一般特征分析方法中，表示变量作用的变量权取为特征向量α的分量α_j，它反映第j个原始变量在这个特征方向上的作用，亦即该变量对矿化特征的表现能力。在本方法中，由于非矿样品的使用，第j个变量还有另一个权β_j，它通过非矿特征反映了该变量对矿化的表现能力。一般说，α_j和β_j不相等，它们之间的一致性反映了该变量对矿和非矿的区分意义大小。因此可用α_j和β_j的乘积或代数和来度量第j个变量的重要性。换句话说，一个重要的变量不仅表现为有较

大的α分量，也应有较大的β分量。

根据两套变量权，每个样品都可求得一对关联度：矿关联度u和非矿关联度v。其中矿样品关联度为：

$$u_{(x)}=\alpha_1 x_1+\alpha_2 x_2+\cdots+\alpha_m x_m$$
$$=X\alpha$$
$$v_{(x)}=\beta_1\widetilde{x}_1+\beta_2\widetilde{x}_2+,\cdots,+\beta_m\widetilde{x}_m$$
$$=\widetilde{X}\beta,$$

这里$\widetilde{x}_j$、$\widetilde{X}$都是取补的含义。非矿样品关联度为：

$$u_{(y)}=\alpha_1 y_1+\alpha_2 y_2+\cdots+\alpha_m y_m$$
$$=Y\alpha$$
$$v_{(y)}=\beta_1\widetilde{y}_1+\beta_2\widetilde{y}_2+\cdots+\beta_m\widetilde{y}_m$$
$$=\widetilde{Y}\beta$$

通过对两类样品关联度的某些特征如秩顺序、平均数和上下界等的对比，可以构造预测样品含矿性的判别标准。例如，将模型样品按关联度标示在u—v平面上（图1），两类样品点大体集中在两个区域，在矿样品（实心点）集中的区域Ⅰ上关联度的特点为u较大，v较小；而在非矿样品（空心点）集中的区域Ⅱ上u较小，v较大。若将Ⅰ、Ⅱ两区域分别看作是矿域和非矿域，则可根据预测样品的u、v值落入各区域情况判断其归属。

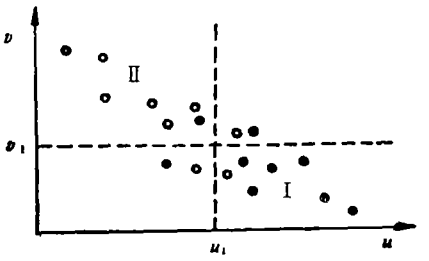

图 1　在u—v上两类样品点的分布

含矿性评价

继续考虑通过样品关联度u、v给出预测样品含矿性概率估计的方法，即建立起关联度与含矿概率之间的关系。为此，首先根据两类模型样品求出矿母体和非矿母体关联度的分布。

在数据矩阵X中，对变量x_j求出它在s个矿样品上取1的频率p_j和相应取0的频率$1-p_j$，于是得到变量x_j取值的一个0—1分布。假设对全部m个变量建立了m个这样的0—1分布，则可在这些分布中使用蒙特卡罗技术进行反复抽样，再对足够多的抽样样品进行统计整理，从而建立起矿母体的样品分布。

具体抽样方法是：设随机抽取的样品为$w=(\delta_1,\ \delta_2,\ \cdots,\ \delta_m)$，$\delta_j$是样品$w$在第$j$个变量上的取值，它究竟是取0还是取1要由抽样结果决定。首先，取(0，1)上均匀分布的随机数r_1，若$r_1\leqslant p_1$则令$\delta_1=1$，否则$\delta_1=0$；然后再取r_2，若$r_2\leqslant p_2$则$\delta_2=1$，否则$\delta_2=0$；直到取r_m确定了δ_m的值。这样即得到一个随机抽样样品w。按此方法，每取一组m个随机数就完成一轮抽样，得到一个随机样品。假如一共进行了N轮抽样，于是得到N个随机样品。

通过抽样得到的样品w，不一定与X中的某个模型样品相同，而对w按变量权α、β求得的一对关联度当然也不一定等于某个模型样品的$u_{(x)}$、$v_{(x)}$值。最终目的是建立起样品关联度与含矿概率之间的关系，因此在每次抽样的同时，就可求出一对关联度，当抽样次数足够大时，抽样样品关联度则形成一个二维分布，它可近似看作是矿母体关联度的理论分布。

依同样的方法，在数据矩阵$\widetilde{Y}$中求出各变量取1的频率q_j和取0的频率$1-q_j$，$(j=1,\ 2,\ \cdots,\ m)$，然后进行抽样，也可得到非矿母体关联度的近似理论分布。

由于理论分布代表了模型样品关联度的特征，同时随机抽样样品包容了与模型样品成矿地质条件相似的所有情况，因此作为预测对比的标准，理论分布要比模型样本更合理。在这种情况下，对预测对象来说，它与

模型的类比就相当于在理论分布中确定它的位置特征。根据这一想法，便可利用关联度理论分布来构造一个方法解决含矿概率估计问题。

假定已通过抽样模拟的方法得到N_1个矿和N_2个非矿样品，每个样品都有一对关联值，则可用$u_{(x)}$、$v_{(x)}$表示矿样品关联度和用$u_{(y)}$、$v_{(y)}$表示非矿样品关联度。把所有样品点标示在$u-v$平面上（图2），实心点为矿样品，空心点为非矿样品。A、B两点分别是两个点群的重心，坐标为（$\overline{u}_{(x)}$，$\overline{v}_{(x)}$）和（$\overline{u}_{(y)}$，$\overline{v}_{(y)}$），其中，

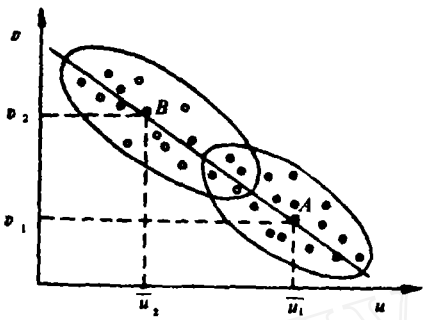

图 2　两类抽样样品点的分布

$$\left.\begin{aligned}\overline{u}_{(x)}&=\frac{1}{N_1}\sum_{i=1}^{N_1}u_{(x)i},\\ \overline{v}_{(x)}&=\frac{1}{N_1}\sum_{i=1}^{N_1}v_{(x)i},\\ \overline{u}_{(y)}&=\frac{1}{N_2}\sum_{i=1}^{N_2}u_{(y)i},\\ \overline{v}_{(y)}&=\frac{1}{N_2}\sum_{i=1}^{N_2}v_{(y)i},\end{aligned}\right\}\qquad(1)$$

通过点A、B做一条直线l，l的方程为：

$$au+bv+c=0,$$

其中，

$$a=\overline{v}_{(y)}-\overline{v}_{(x)},$$

$$b=\overline{u}_{(x)}-\overline{u}_{(y)},$$

$$c=\overline{u}_{(y)}\overline{v}_{(x)}-\overline{u}_{(x)}\overline{v}_{(y)}.$$

将平面$u-v$上所有样品点投影到直线l上，那么问题就可以转化到一维空间去研究。设第j个抽样样品（不论是矿还是非矿）的一对关联度为(u_j, v_j)，作变换：

$$e_j=u_j-a(au_j+bv_j+c)/(a^2+b^2)\qquad(2)$$

e_j兼顾u_j、v_j两个关联度值，是对第j个样品的综合标度。把e_j叫作第j个样品的对象权，它是样品在l上投影的横坐标。容易验证，A、B两点的对象权即为$\overline{u}_{(x)}$和$\overline{u}_{(y)}$。

在实际计算中，抽样样品的对象权可在抽样中同时求得。如果采用模型样本的频率来代替理论分布的概率，那么A、B的坐标则不依赖于抽样结果，因而直线l的方程可在抽样之前确定，于是每抽得一个随机样品，便可根据它的u、v值按（2）式求得e。此时（1）式中的4个方程式分别变为：

$$\left.\begin{aligned}\overline{u}_x&=\sum_{j=1}^{m}p_j\alpha_j,\\ \overline{v}_x&=\sum_{j=1}^{m}(1-p_j)\beta_j,\\ \overline{u}_y&=\sum_{j=1}^{m}(1-q_j)\alpha_j,\\ \overline{v}_y&=\sum_{j=1}^{m}q_j\beta_j\end{aligned}\right\}\qquad(3)$$

需要说明，随机抽样得到的矿和非矿样品要分别统计，以便建立两类样品对象权各自的分布。假定在两类中，抽样分别为N_1、N_2次，各得矿和非矿对象权N_1和N_2个，对两类e值分别进行整理，通过模拟方法做出两条概率密度曲线。下面根据两条曲线的关系来构造矿和非矿的分类判别准则。

图3是两类样品对象权概率分布示意图，其中h_1是矿样品对象权的分布密度，h_2是非矿样品对象权的分布密度。取两条曲线h_1、h_2交点所对应的对象权$e=e_0$作为判据，e_0是方程$h_1=h_2$的解，则判别准则为：对任一预测样品，若求得对象权$e\geqslant e_0$则将其判为矿，若$e<e_0$则判为非矿。由图3中可见，当根据$e<e_0$将预测样品判为非矿时发生错误

的概率正是阴影部分的面积，这种错误把本应是矿的样品判为非矿。因此，判别一个预

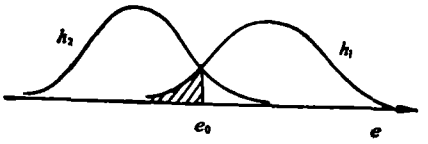

图 3　两类抽样样品对象权分布示意图

测样品为非矿时判错率可用下式表示：

$$p' = \int_0^{e_0} h_1(x)dx,$$

类似地，当将预测样品判为矿时判错率为：

$$p'' = 1 - \int_0^{e_0} h_2(x)dx$$

具体求h_1和h_2的表达式，通常采用已知概率分布模拟的方法。根据概率论的中心极限定理，随机变量e渐近服从正态分布，证明过程略去。于是，由抽样样品对象权求出各类均值：

$$\bar{e}_{(x)} = \frac{1}{N_1}\sum_{i=1}^{N_1} e_{(x)i}$$

$$\bar{e}_{(y)} = \frac{1}{N_2}\sum_{i=1}^{N_2} e_{(y)i}$$

和标准差：

$$\sigma_{(x)} = \sqrt{\frac{1}{N_1}\sum_{i=1}^{N_1}(e_{(x)i} - \bar{e}_{(x)})^2}$$

$$\sigma_{(y)} = \sqrt{\frac{1}{N_2}\sum_{i=1}^{N_2}(e_{(y)i} - \bar{e}_{(y)})^2},$$

可直接写出：

$$h_1(e) = \frac{1}{\sqrt{2\pi}\sigma_{(x)}}\exp\left[\frac{(e-\bar{e}_{(x)})^2}{2\sigma^2_{(x)}}\right],$$

$$h_2(e) = \frac{1}{\sqrt{2\pi}\sigma_{(y)}}\exp\left[\frac{(e-\bar{e}_{(y)})^2}{2\sigma^2_{(y)}}\right]$$

而判据由公式：

$$e_0 = \frac{\bar{e}_{(x)}\sigma^2_{(y)}\ln\sigma_{(x)} - \bar{e}_{(y)}\sigma^2_{(x)}\ln\sigma_{(y)} \pm (\bar{e}_{(x)} - \bar{e}_{(y)})\sigma_{(x)}\sigma_{(y)}\sqrt{\ln\sigma_{(x)}\ln\sigma_{(y)}}}{\sigma^2_{(y)}\ln\sigma_{(x)} - \sigma^2_{(x)}\ln\sigma_{(y)}} \qquad (4)$$

有两个解，去掉增根即得e_0。

下面讨论两个特征方向α和β之间的关系对预测结果的影响。设α和β的夹角为φ，$\varphi = \arccos\alpha'\beta$。由直线$l$的取法可知，它通过两个样品点群的重心$A$和$B$。所有样品点分布在$l$的两侧，它们与$l$的靠近程度依赖于$\varphi$，当$\varphi=0$时，由$\alpha=\beta$有$\alpha_j=\beta_j$，于是，推出$v_{(x)} = \sum_{j=1}^{m}\alpha_j - u_{(x)}$和$v_{(y)} = \sum_{j=1}^{m}\beta_j - u_{(y)} = \sum_{j=1}^{m}\alpha_j - u_{(y)}$，即$u$和$v$有函数关系，故全部样品点都落在$l$上。此时公式（2）变成：

$$e_j = u_j$$

当$\varphi\neq 0$时，样品点及其投影e的离散程度随φ变大，h_1、h_2的曲线变缓，此时两个分布的均值不变而方差增大　影响到图3中阴影面积增大。因此，α、β的相似程度影响判错率，从而影响到模型的可靠性。

至此，矿床模型的建立已告完成，对于研究一般问题，样品含矿性评价做到分类这一步也足够了。下面考虑的内容，是根据矿产资源评价中对预测成果的某些具体要求来讨论的，可把它看作是这一方法在资源评价中解决定位预测问题的应用。

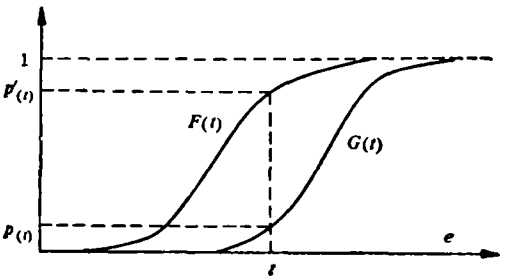

图 4　两类对象权分布函数示意图

做出 (0, 1) 上h_1和h_2的概率分布函数:

$$F(t)=\int_0^t h_1(x)dx,$$

$$G(t)=\int_0^t h_2(x)dx,$$

两个函数的图形(图4)，其中右边曲线$F(t)$表示矿样品关联度$e\leqslant t$时的概率。图4中$e=t$处对应概率为$p=p_t$，可解释为：在所有矿样品中对象权小于 t 的概率不超过p_t。分布函数有如下性质:

① 非负;

② 有界。当$t\to 0$时，$F(t)\to 0$; 当$t\to -\frac{c}{a}$时，$F(t)\to 1$;

③ $F(t)$在 t 较大或较小时变化平稳，在 t 取中间值时变化幅度较大。

由上述性质可知，p_t满足“具备有利条件越多的样品成矿可能性越大”的基本假设，符合对样品标度与含矿性进行关联的要求，并与找矿经验吻合。再用$F(t)=p_t$来构造对象权$e=t$时预测样品的含矿概率，它表示对象权小于 t 的预测样品含矿概率不超过p_t。

用类似的方法也可构造非矿概率函数$G(t)$(图4左)，它与$F(t)$一样满足作为概率估计的条件，只是需要注意它的意义：与$G(t)$相应的p_t'表示对象权$e=t$时样品不是非矿的概率。

前已述及，在模型中引入非矿样品的目的在于造成对样品的双重度量，从矿和非矿两个方面来研究样品的标度特征。因此，构造预测样品的含矿概率用乘积:

$$\psi(t)=F(t)\cdot G(t)=p_t\cdot p_t' \quad (5)$$

来表示。函数$\psi(t)$性质与$F(t)$相差不多，在两类样品特征α和β独立求出的情况下，乘积$p_t\cdot p_t'$可看作对样品是矿且不是非矿的双重计量。

计算实例

为建立某类型矿床的资源评价模型，在研究区内选取14个网格单元作为模型样品，其中含矿单元6个，不含矿单元8个。使用变量12个，原始赋值矩阵为$\boldsymbol{X}=(x_{ij})_{8\times 12}$和$\boldsymbol{Y}=(y_{ij})_{8\times 12}$，

$$\boldsymbol{X}=\begin{pmatrix}1&0&1&1&1&1&1&1&1&1&1&0\\1&0&1&1&0&0&1&0&0&1&1&1\\0&0&1&1&1&1&0&0&1&1&1&1\\1&1&1&1&1&1&0&1&1&0&1&0\\0&1&1&0&1&1&0&1&1&1&1&0\\1&0&1&0&1&0&0&1&0&1&0&1\end{pmatrix}$$

$$\boldsymbol{Y}=\begin{pmatrix}1&1&1&0&0&0&1&1&1&0&0&0\\0&0&0&1&1&1&0&0&0&1&0&1\\0&0&1&1&0&0&0&1&0&0&1&1\\0&0&1&1&0&0&1&1&1&1&0&0\\0&0&1&0&1&1&0&0&1&0&0&1\\0&0&0&0&0&1&0&0&0&0&1&0\\0&0&0&0&0&1&0&0&1&1&0&1\\1&0&1&1&0&0&0&1&0&0&0&0\end{pmatrix}$$

取$\boldsymbol{Y}$的补矩阵$\widetilde{\boldsymbol{Y}}$，

$$\widetilde{\boldsymbol{Y}}=\begin{pmatrix}0&0&0&1&1&1&0&0&0&1&1&1\\1&1&1&0&0&0&1&1&1&0&1&0\\1&1&0&0&1&1&1&0&1&1&0&0\\1&1&0&0&1&1&0&0&0&0&1&1\\1&1&0&1&0&0&1&1&0&1&1&0\\1&1&1&1&1&0&1&1&1&1&0&1\\1&1&1&1&1&0&1&1&0&0&1&0\\0&1&0&0&1&1&1&0&1&1&1&1\end{pmatrix}$$

$\boldsymbol{X}$和$\widetilde{\boldsymbol{Y}}$的变量匹配矩阵为

$$X'X=\begin{pmatrix}
4&1&4&3&3&2&2&3&2&3&3&2\\
1&2&2&1&2&2&0&2&2&1&2&0\\
4&2&6&4&5&4&2&4&4&5&5&3\\
3&1&4&4&3&3&2&2&3&3&4&2\\
3&2&5&3&5&4&1&4&4&4&4&2\\
2&2&4&3&4&4&1&3&4&3&4&1\\
2&0&2&2&1&1&2&1&1&2&2&1\\
3&2&4&2&4&3&1&4&3&3&3&1\\
2&2&4&3&4&4&1&3&4&3&4&1\\
3&1&5&3&4&3&2&3&3&5&4&3\\
3&2&5&4&4&4&2&3&4&4&5&2\\
2&0&3&2&2&1&1&1&1&3&2&3
\end{pmatrix}$$

$$\widetilde{Y}'\widetilde{Y}=\begin{pmatrix}
6&6&3&3&4&2&5&4&3&3&4&2\\
6&7&3&3&5&3&6&4&4&4&5&3\\
3&3&3&2&2&0&3&3&2&1&2&1\\
3&3&2&4&3&1&3&3&1&3&3&2\\
4&5&2&3&6&4&4&2&3&4&4&4\\
2&3&0&1&4&4&2&0&2&3&3&3\\
5&6&3&3&4&2&6&4&4&4&4&2\\
4&4&3&3&2&0&4&4&2&2&3&1\\
3&4&2&1&3&2&4&2&4&3&2&2\\
3&4&1&3&4&3&4&2&3&5&3&3\\
4&5&2&3&4&3&4&3&2&3&6&3\\
2&3&1&2&4&3&2&1&2&3&3&4
\end{pmatrix}$$

求出矩阵$X'X$的最大特征值$\lambda_1=35.585$，相应的特征向量为α，$\alpha=(0.265,\ 0.144,\ 0.405,0.285,0.350,0.299,0.139,\ 0.280,\ 0.299,\ 0.329,\ 0.354,\ 0.176)'$，而$\widetilde{Y}'\widetilde{Y}$的最大特征值$\mu_1=39.306$，相应的特征向量为$\beta$，$\beta=(0.345,0.405,0.187,0.230,0.336,0.202,0.358,0.243,0.241,\ 0.284,\ 0.318,\ 0.220)'$，$\alpha$和$\beta$是含矿和不含矿单元的两个主要特征方向，把它们的分量取为两类单元相应的变量权。根据这两组变量权，分别计算含矿单元的两种关联度$u_{(x)}=\sum_{j=1}^{m}\alpha_j x_j$和$v_{(x)}=\sum_{j=1}^{m}\beta_j\ \widetilde{x}_j$，全部6个含矿单元的计算结果列在表1中。

含矿单元关联度　　表 1

单元号	1	2	3	4	5	6
$u_{(x)}$	3.006	1.954	2.497	2.681	2.460	1.805
$v_{(x)}$	0.624	1.428	1.351	0.862	1.153	1.754

类似地，用公式$u_{(y)}=\sum_{j=1}^{m}\alpha_j y_j$和$v_{(y)}=\sum_{j=1}^{m}\beta_j\ \widetilde{y}_j$亦可求出全部不含矿单元的两种关联度(表2)。

按两种关联度在平面u—v上表示全部单元点(图5)，可以看出含矿单元点和不含矿单元点分布明显不同。含矿单元点的位置靠上，不含矿单元点靠下，这些点大体排列在一个狭窄的带形区域内。

现有一预测单元z，其取值为：

$z=(1\ 0\ 1\ 1\ 1\ 0\ 0\ 0\ 1\ 1\ 1\ 0)'$，

求得$u_z=2.287$，$v_z=1.428$，此单元在图5中用星号表示。显然，预测样品z按关联度落在了含矿单元的分布范围内，因此可以初步考虑将此单元归入含矿类。

为进一步求含矿概率，对矩阵X和$\widetilde{Y}$统计变量取1值的频率，记为$p_{(x)}$和$p_{(y)}$，列在表3中。

根据$p_{(x)}$和$p_{(y)}$，采用蒙特卡罗方法抽样999轮，得到999个含矿的和999个不含矿

不含矿单元关联度　　表 2

单元号	7	8	9	10	11	12	13	14
$u_{(y)}$	1.533	1.439	1.499	1.737	1.529	0.653	1.103	1.235
$v_{(y)}$	1.590	2.097	2.172	1.825	2.182	2.849	2.422	2.364

两类单元状态1频率表　　表 3

变量号	1	2	3	4	5	6	7	8	9	10	11	12
$p_{(x)}$	0.667	0.333	1.000	0.667	0.833	0.667	0.333	0.667	0.667	0.833	0.833	0.500
$p_{(y)}$	0.750	0.875	0.375	0.500	0.750	0.500	0.750	0.500	0.500	0.625	0.750	0.500

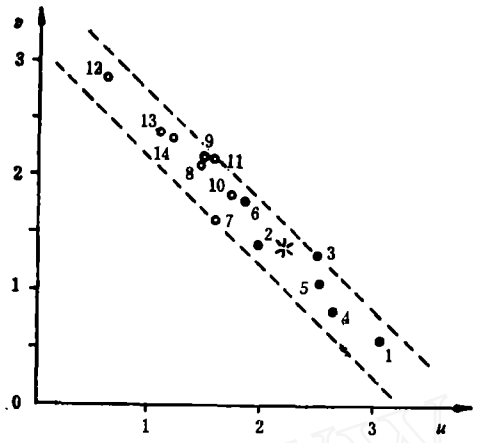

图 5　模型单元关联度示意图

的随机单元。将这些单元分别计算关联度（类似模型单元的作法），也可以将它们点到$u-v$平面上。按公式（3）求得两类单元的重心A和B的坐标分别为(2.400，1.195)和（1.341，2.188)。过A、B点作直线l，方程为：

$$0.993u+1.059v-3.648=0,$$

将所有抽样点投影到l上，变换结果得到两类单元点在一维空间上的数值标度——对象权e。抽样单元中对象权e的最大值为3.490，最小值为0.037，据此将e的取值区间分为10组，间距0.3453，则可作出两类样品关联度的频数(f)和频率(p)表（表4)。

抽样单元频数频率表　　表 4

分组号	1	2	3	4	5	6	7	8	9	10
组中值	0.210	0.555	0.900	0.1246	1.591	1.936	2.282	2.627	2.972	3.318
$f_{(x)}$	0	0	0	19	63	154	321	271	121	50
$f_{(y)}$	51	73	184	270	203	126	60	20	12	0
$p_{(x)}$	0.000	0.000	0.000	0.019	0.063	0.154	0.321	0.271	0.121	0.050
$p_{(y)}$	0.051	0.073	0.184	0.270	0.203	0.126	0.060	0.020	0.012	0.000

根据含矿单元对象权$e_{(x)}$和不含矿单元对象权$e_{(y)}$的频率所做的直方图如图6所示，两个图形均呈单峰、对称形态。用正态分布来模拟对象权的概率分布，为此求出$e_{(x)}$、$e_{(y)}$的均值和标准差，$\bar{e}_{(x)}=2.398$，$\sigma_{(x)}=0.449$；$\bar{e}_{(y)}=1.344$，$\sigma_{(y)}=0.562$。正态分布概率密度函数分别为：

$$h_1=\frac{1}{0.449\sqrt{2\pi}}\exp\left[\frac{(e-2.398)^2}{2\times0.449^2}\right],$$

$$h_2=\frac{1}{0.562\sqrt{2\pi}}\exp\left[\frac{(e-1.344)^2}{2\times0.562^2}\right]$$

两条曲线的形状及位置关系见图7，相应的累积概率曲线见图8。其中判据是据公式（4）求出的 $e_0=1.977$。由预测单元z的关联度u_z和v_z求得对象权$e_z=2.223$，$e_z>e_0$，故可认为该预测单元为含矿单元。

再用公式（5）计算z的含矿率，为此求出$F(e_z)=\int_0^{2.223}h_1(x)\,dx=0.348$和$G(e_z)=\int_0^{2.223}h_2(x)\,dx=0.941$，于是$\psi(e_z)=F(e_z)G(e_z)=0.327$，由此得出结

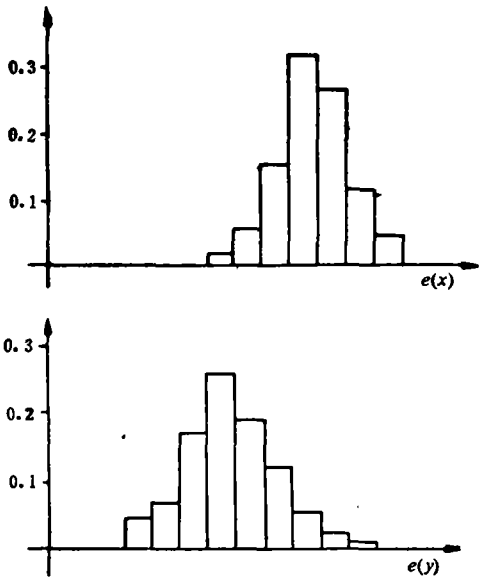

图 6 两类单元对象权的频率直方图

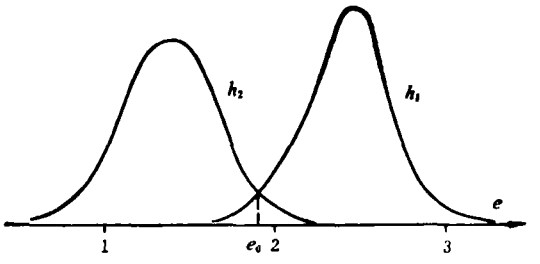

图 7 两类单元对象权概率密度曲线

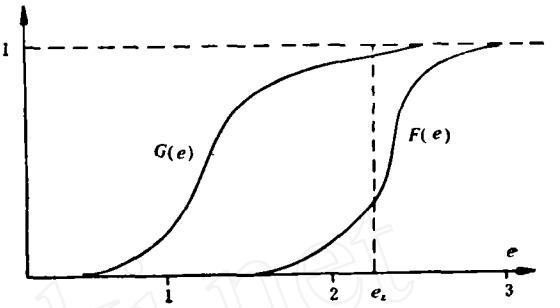

图 8 两类单元对象权分布函数曲线

论，该预测单元的含矿概率为0.327。

本文承夏立显研究员审阅并提出许多有益的意见，特此致谢。

参 考 文 献

[1] 周光亚等，《多元统计分析》，地质出版社，1981年

[2] 杨永华，长春地质学院学报，1988，第1期

A Double Analysis Method for Mineral Resources Assessment

Yang Yonghua

An improved characteristic analysis method is discussed in this paper. Taking a group of nonmineralized samples to form a contrast, the model built manifests the characteristics of the ores from two opposite angles. The forms for comprehensively scaling samples by using the degree of correlation of mineralized and nonmineralized samples and the significance of establishing the theoretical distribution of target weighting are also discussed. An approximate statistical method has been developed to realize the discrimination, classification and ore-bearing property estimation of samples by their contrast.