

基于统计聚类分析的新疆雪米斯坦地区 地球化学元素组合研究

许静¹⁾, 单广宁²⁾

1) 对外经济贸易大学统计学院, 北京, 100029; 2) 中国核工业集团公司地质矿产事业部, 北京, 100013

雪米斯坦火山岩带位于新疆西准地区, 是一条近东西向展布的古生代火山岩带。带内岩浆演化系列完整, 从基性到酸性各类火山熔岩和火山碎屑岩均有发育, 厚度巨大, 是热液型铀-多金属矿成矿的有利地区。“十一五”以来, 该区一直被列为我国铀矿勘查的重点地区, 投入了大量的地、物、化、遥等工作, 尤其是系统开展了 1:50000 区域化探测量, 圈定出一批地球化学异常^①。前人成矿规律研究成果表明, 区内成矿热液是多期次的, 每期热液性质不同, 成矿元素也存在差异。因此, 有必要在该区开展地球化学数据聚类分析, 确定与成矿有关的地球化学元素组合异常, 为在雪米斯坦地区进一步部署铀-多金属找矿工作提供依据。

1 聚类分析

1.1 聚类分析的原理

在多元统计中, 聚类分析认为研究的对象(样本或变量)之间存在着不同程度的相似性, 根据一些能够度量对象之间相似程度的统计量, 可划分对象的类型。把一些相似程度较大的对象聚合为一类, 把另外一些相似程度较大的对象聚合为另一类; 关系密切的聚合到一个小的分类单位, 关系疏远的聚合到一个大的分类单位; 直到把所有的对象都聚合完毕, 形成一个由小到大的分类系统, 将亲疏关系表示出来(何晓群, 2008)。

1.2 系统聚类法

系统聚类分析, 首先将所研究的每个对象自己看作一个类, 计算相似矩阵, 反映相互之间的接近程度, 将最相近的优先合并为一类; 然后进一步计算类与类之间的距离, 再合并相近的类, 直至将所有对象合并为一个大类。系统聚类的过程实际上给

出了从最细的分类到最粗的分类之间的所有分类结果; 最后, 根据研究需要, 可以将对象分为若干类, 即选择聚类过程中的一个分类结果。

设有 n 个样品, 每个样品测 p 项指标(变量) X_1, X_2, \dots, X_p , 所得数据构成矩阵

$$\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} \quad (1)$$

其中, x_{ki} 表示第 k 个样品的第 i 个指标的测试结果。

变量间的相似性可以用相关性进行评价, 相关系数是最常用的评价指标。变量 X_i 与 X_j 的相关系数定义为

$$r_{ij} = \frac{\sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{\sqrt{\sum_{k=1}^n (x_{ki} - \bar{x}_i)^2 \sum_{k=1}^n (x_{kj} - \bar{x}_j)^2}} \quad (2)$$

其中, $\bar{x}_i = \frac{1}{n} \sum_{k=1}^n x_{ki}$, $i = 1, 2, \dots, n$, 是数据(1)中第 i 个指标的样本均值。

类与类之间的距离采用类间平均距离, 设 G_p 和 G_q 是分别含有 n_p 个和 n_q 个元素的两类, 其类间

平均距离为 D_{pq} , 定义为 $D_{pq}^2 = \frac{1}{n_p n_q} \sum_{X_i \in G_p} \sum_{X_j \in G_q} d_{ij}^2$, 即

这两类元素两两之间距离平方的平均值, 其中, $d_{ij}^2 = \sum_{k=1}^n (x_{ki} - x_{kj})^2$ 表示元素 X_i 与 X_j 间的平方欧氏距离。

收稿日期: 2015-01-30; 改回日期: 2015-03-13; 责任编辑: 章雨旭。
作者简介: 许静, 女, 1977 年生。博士, 副教授, 统计学。Email: fouringx@163.com。

2 数据源

本次研究使用的地球化学数据来源于核工业二一六大队以及核工业二三〇、二七〇研究所有关项目组在该区开展的 1:50000 岩屑测量的获得的 1136 组化探数据。虽然每个项目对单个样品均测量了 10 多种元素,但因为每个项目测试的元素种类不一样,仅有 U、Be、Au、Cu、Pb、Zn、Li 和 Mn 8 种元素覆盖全区,所以仅对这 8 种元素进行了聚类分析,设为变量 X_1, X_2, \dots, X_8 。

因为这 8 种元素在岩石中的含量差别较大(从 10^{-2} 到 10^{-7}),需要对数据进行标准化。以 X_i 为例,

X_i 的观测 x_{ki} 标准化之后为 $x_{ki}^* = \frac{x_{ki} - \bar{x}_i}{s_i}$ 。其中, \bar{x}_i

是样本均值; $s_i = \sqrt{\frac{1}{n-1} \sum_{k=1}^n (x_{ki} - \bar{x}_i)^2}$, 是样本的标准差。

3 聚类分析结果与讨论

3.1 聚类分析结果

应用统计软件 SPSS 对雪米斯坦地区 1136 个化探样品的 8 种元素的测试结果先进行标准化,然后进行聚类分析。根据相关系数公式(2)得到相关系数矩阵(表 1),进而得到聚类树状图(图 1)。

表 1 雪米斯坦地区地球化学元素相关系数矩阵

相关系数	U	Be	Au	Cu	Pb	Zn	Li	Mn
U	1.000	.373	-.037	-.034	.180	.112	-.187	-.244
Be	.373	1.000	-.051	-.032	.076	.063	-.253	-.279
Au	-.037	-.051	1.000	.137	.010	-.003	.103	.037
Cu	-.034	-.032	.137	1.000	.012	.038	.082	.082
Pb	.180	.076	.010	.012	1.000	.587	-.091	-.043
Zn	.112	.063	-.003	.038	.587	1.000	.119	.268
Li	-.187	-.253	.103	.082	-.091	.119	1.000	.522
Mn	-.244	-.279	.037	.082	-.043	.268	.522	1.000

相关系数代表了元素两两间相关性的强弱。由表 1 可知, Pb 和 Zn 的相关系数为 0.587, 是相关性最强的一组元素; Li 和 Mn 相关系数为 0.522, 是相关性较强的一组元素; U 和 Be 相关系数为 0.373, 是具有一定相关性的一组元素。这 8 种元素的相关性也体现在聚类分析的结果中。

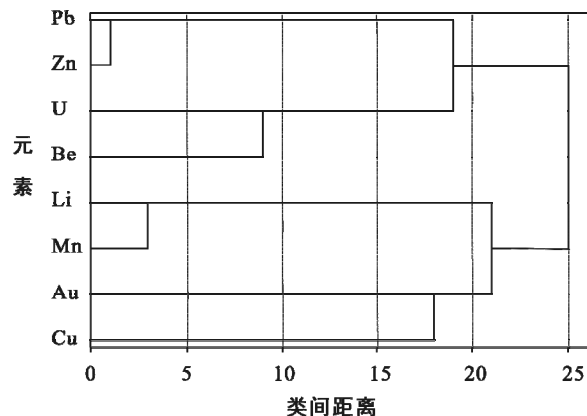


图 1 新疆雪米斯坦地区地球化学元素聚类树状图

由图 1 可知, 第一个层次的聚类将元素分为 4 组。其中, Pb 和 Zn 类间距离最短, 首先聚为一类; Li 和 Mn 间的类间距离是新的类中最短的, 聚为一类; 依次, U 和 Be 聚为一类, Au 和 Cu 聚为一类。再计算这四类之间的距离, 又聚为两类, 一类是 Pb—Zn—U—Be 元素组合, 另一类是 Cu—Au—Li—Mn 元素组合。

3.2 结果讨论

地球化学元素组合取决于元素的地球化学属性, 与地质构造演化和热液成矿作用有关。在雪米斯坦地区, 杨庄等花岗斑岩中富含 Be、U、Mo、Pb、Zn 和 V 等元素, 后期又叠加了富含 U、Mo、Be 等元素的多期热液活动改造, 形成了 U—Pb—Zn—Be 元素组合; 在除花岗斑岩等岩体以外地区, 中基性、中酸性熔岩与碎屑岩交替产出, 富含 Cu、Co、Cr、Ni、Sb、Mn 和 Sn 等元素, 不同岩性元素含量有一定的差异, 在后期碱性热液的改造下形成 Cu—Mn—Au—Li 元素组合。

本次研究确定的两个元素组合, 与前人^①在雪米斯坦地区开展化探时确定的 Be—U—Mo—Pb—Zn—V 和 Cu—Co—Cr—Ni—Sb—Mn—Sn 两个元素组合具有明显的一致性。通过系统圈定这两个元素组合的异常, 可作为在该区开展铀、铍等矿产资源预测评价的依据。

注 释 / Note

①周剑, 李彦龙, 陈奋雄, 等. 2011. 新疆和布克赛尔县白杨河地区 1:5 万铀矿区域地质调查, 核工业二一六大队地质调查报告, 102~117.

参 考 文 献 / References

何晓群. 2008. 多元统计分析(第 2 版), 中国人民大学出版社, 57~95.