

样品集合分类方法及其在 化探异常评价中的应用

胡 坤

提 要 样品集合分类法是一种新的分类方法。该方法的最大优点是参与分类的各个样品不再是单个的样品点,而是已经形成的一个点群。并以离差平方和法为数学模型,以关门山铅锌矿区为例子,说明样品集合分类法的应用。

关键词 样品集合 样品集合分类 离差平方和

一、样品集合分类方法及其优点

所谓样品集合,就是具相同或相似特征的某一类样品的总体。至于样品个数的多少,可以不定,但一个样品不能称之为样品集合。

地质上的许多分类问题,都可以归为样品集合分类。例如:在化探异常分类中,假设某区有 g 个异常,每个异常被 $n_i(i=1,2,\dots,g)$ 个样品控制,每个样品分析了 m 个指标(不妨设这 m 个指标在各个异常中服从 m 维正态分布),要对这 g 个异常定量分类。通常,地质工作者会很容易想到用聚类、因子、对应等多元分析。我们知道,一般的多元分析只适用于单个样品的分群,对这种一群群样品的分类,它们是无能为力的。因为一般的多元分析不能让每个异常的 $n_i(i=1,2,\dots,g)$ 个样品同时参与分类,如果让每个异常的 $n_i(i=1,2,\dots,g)$ 个样品都参与分类,由于在分类过程中它们不能保证让各个异常体的样品不被“打散”,结果会使某一异常的样品分到其它各个异常中,致使分类结果混乱,达不到分类的目的。为此,只能求出每个异常的重心样品,用它代替各异常去参与分类,以达到分类之目的。然而,这就损失了很多信息。从统计学的角度看,样品重心只是正态分布的均值向量,由正态分布密度函数的几何意义可知,均值只能控制正态分布密度函数图形的位置,而不能控制其形状,用均值代替样品集合参与分类,会使均值相同,离差不同的样品集合分为一类,从而得出错误结论。为了使分类结果正确,就必须利用样品集合分类法。

什么是样品集合分类呢?可以这样定义:如果我们把上述 g 个异常中的每个异常的 $n_i(i=1,2,\dots,g)$ 个样品看成一个整体,让它们一同参与分类(即在分类过程中都参加,但不“打散”各样品集合的样品),这就叫样品集合分类方法。与一般的分类方法相比,参与分类的各个样品不再是单个的样品点,而是已经形成的一个点群。事实上,当各个样品集合中所包含的样品个数均为1时,样品集合分类法就成为一般的样品分类法。

样品集合分类方法的最大优点是:它能充分利用原始资料所提供的信息,最大限度地反映各样品集合之间的亲近程度。

样品集合分类方法有许多数学模型,比如象典型变量分析^①(不同于典型相关),离差平方和法等,下面我们以离差平方和法为例,来说明样品集合分类方法的应用

二、离差平方和法数学模型简介^②

离差平方和法事实上也是一种 Q 型聚类分析方法。在用于单个样品分类时,其优点在于:可以消除用距离系数作分群时可能出现的反转现象。但离差平方和法还有一个被人们忽视的、也是很重要的优点,那就是它可以用于样品集合分类。它的分类准则是:如果分类正确,同类样品之间的离差平方和就应当较小,类与类之间的离差平方和就应当较大。

离差平方和是这样定义的,对于已经形成的各个点群,算出它的重心,再算出这个点群中所有各点到重心的距离平方和,叫做这个点群的离差平方和。如果将两个点群合并到一起得一新点群,从新点群的离差平方和中减去原来两个点群离差平方和所得的差,叫离差平方和的增量。每一阶段都将离差平方和增量最小的两个点群合并到一起。离差平方和数学模型描述如下:

设有 g 个点群,其中第 K 个($K=1,2,\dots,g$)点群含 n_k 个样品点,记为

$$\vec{X}_{ik}, i=1,2,\dots,n_k; k=1,2,\dots,g$$

第 k 个点群的重心为

$$\vec{X}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} \vec{X}_{ik} \quad k=1,2,\dots,g$$

这点群的组内离差平方和为

$$\begin{aligned} E_k &= \sum_{i=1}^{n_k} \|\vec{X}_{ik} - \vec{X}_k\|^2 = \sum_{i=1}^{n_k} (\vec{X}_{ik} - \vec{X}_k)' (\vec{X}_{ik} - \vec{X}_k) \\ &= \sum_{i=1}^{n_k} \|\vec{X}_{ik}\|^2 - n_k \|\vec{X}_k\|^2 \end{aligned}$$

若将第 p, q 两个点群合并到一起而成新点群 t , 则新点群的重心为

$$\vec{X}_t = \frac{1}{n_t} (n_p \vec{X}_p + n_q \vec{X}_q), n_t = n_p + n_q$$

它的离差平方和为

$$E_t = \sum_{i=1}^{n_p} \|\vec{X}_{ip}\|^2 + \sum_{i=1}^{n_q} \|\vec{X}_{iq}\|^2 - n_t \|\vec{X}_t\|^2$$

合并后所得的离差平方和增量为

$$\begin{aligned} \Delta E_{pq} &= E_t - E_p - E_q \\ &= n_p \|\vec{X}_p\|^2 + n_q \|\vec{X}_q\|^2 - n_t \|\vec{X}_t\|^2 \end{aligned}$$

① 周光亚,石立显,多维标度法及其应用,1986

② 长春地质学院数学地质教研室,多元统计分析方法在地质上的应用,1983

设已将点群 p、q 合并成新点群 t, 对于另外的任一点群 r, 可以证明^[3]增量 ΔE_r 的刷新公式为

$$\Delta E_{tr} = \frac{1}{n_t + n_r} [(n_p + n_r) \Delta E_{pr} + (n_q + n_r) \Delta E_{qr} - n_r \Delta E_{pq}]$$

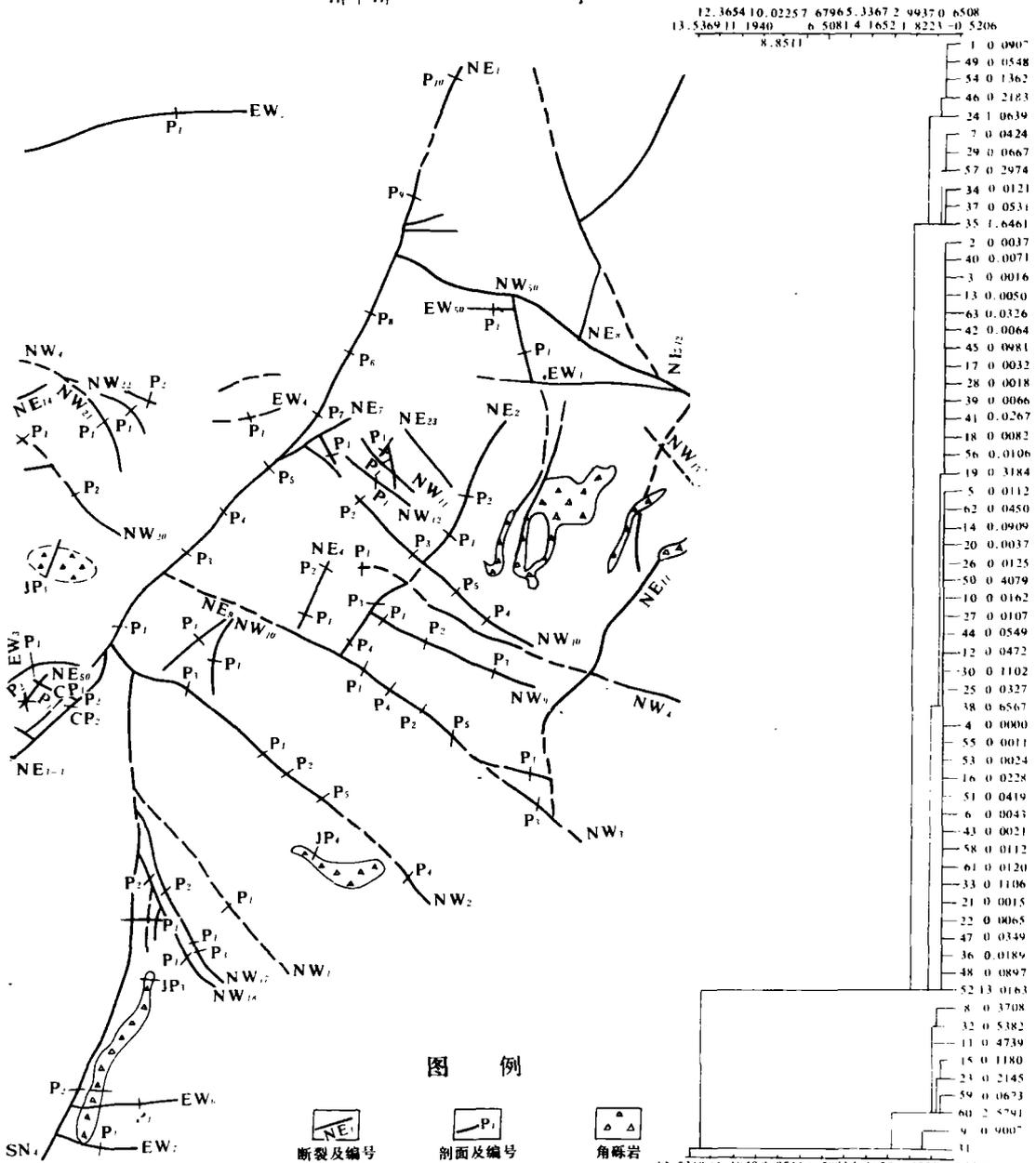


图1 辽宁关门山铅锌矿区构造化探采样位置图

Fig. 1 Location map of geochemical samples along fractures in Guanmenshan

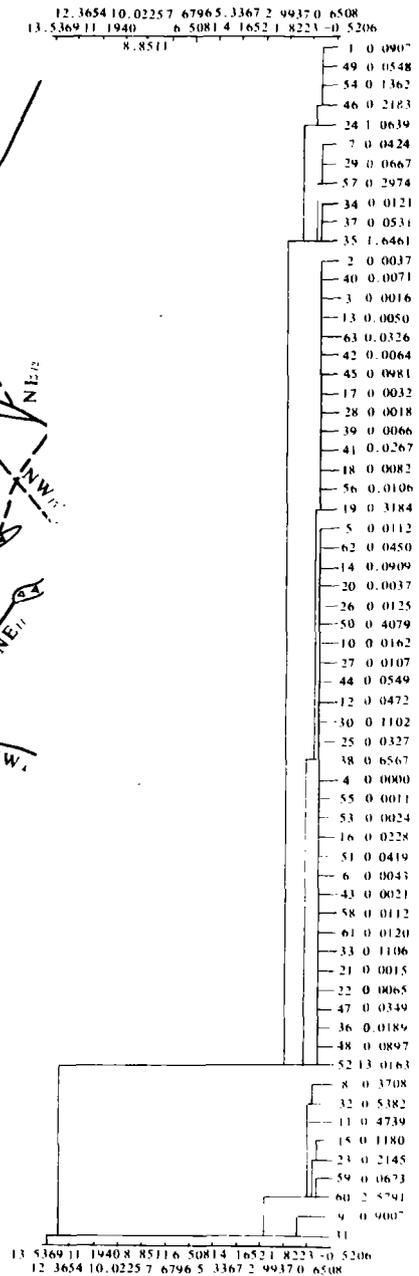


图2 离差平方和法分类谱系图

Fig. 2 Dispersion square sum hierarchy diagram

在给样品集合分群时,我们可用增量刷新公式来逐步实现。

表1 样品集合一览表

Table 1 Sample group schedule

编号	剖面名称	样品数	属性	编号	剖面名称	样品数	属性
1	EW ₃ P ₁	22	已	33	NW ₁₇ P ₁	39	未
2	EW ₄ P ₁	23	未	34	NW ₁₇ P ₂	11	未
3	EW ₄ P ₂	9	未	35	NW ₁₇ P ₃	16	未
4	EW ₅₀ P ₁	7	未	36	NW ₁₈ P ₁	18	未
5	EW ₆ P ₁	8	未	37	NW ₁₈ P ₂	11	未
6	EW ₇ P ₁	6	未	38	NW ₁₉ P ₁	10	未
7	NW ₁₀ P ₁	8	未	39	NW ₁ P ₁	10	未
8	NE ₁ P ₁	28	已	40	NW ₂₀ P ₁	8	未
9	NE ₁ P ₂	30	已	41	NW ₂₀ P ₂	16	未
10	NE ₁ P ₃	32	已	42	NW ₂₁ P ₁	28	未
11	NE ₁ P ₄	11	已	43	NW ₂₂ P ₁	17	未
12	NW ₁ P ₅	17	未	44	NW ₂ P ₁	19	未
13	NE ₁ P ₅	7	未	45	NW ₂ P ₂	15	未
14	NE ₁ P ₇	8	未	46	NW ₂ P ₃	15	未
15	NE ₁ P ₈	16	未	47	NW ₂ P ₄	22	未
16	NE ₂₃ P ₁	9	未	48	NW ₂ P ₅	10	未
17	NE ₂ P ₂	11	未	49	NW ₃ P ₁	30	未
18	NE ₂ P ₂	19	未	50	NW ₃ P ₂	13	未
19	NE ₂ P ₃	8	未	51	NW ₃ P ₃	22	未
20	NE ₂ P ₄	5	未	52	NW ₃ P ₄	14	未
21	NE ₄ P ₁	10	未	53	NW ₃ P ₅	9	未
22	NE ₄ P ₂	6	未	54	NW ₄ P ₁	26	未
23	NW ₅₀ P ₁	14	已	55	NW ₅ P ₁	7	未
24	NE ₅₀ P ₂	21	已	56	NW ₉ P ₁	8	未
25	NE ₇ P ₁	24	未	57	NW ₉ P ₂	7	未
26	NE ₉ P ₁	4	未	58	NW ₉ P ₃	10	未
27	NW ₁₀ P ₁	22	未	59	P ₁	19	已
28	NW ₁₀ P ₂	20	未	60	P ₂	12	已
29	NW ₁₀ P ₃	18	未	61	P ₁	30	未
30	NW ₁₀ P ₄	13	未	62	SN ₄ P ₁	23	未
31	NW ₁₀ P ₅	9	未	63	SN ₄ P ₂	9	未
32	NW ₁₂ P ₁	12	未				

三、离差平方和法应用实例

我们把离差平方和法应用于辽宁关门山铅锌矿区。矿区地质状况简介如下(见图1):

矿区面积约57平方公里,主要断裂为两组方向:NE和NW,其中最大的断裂为NE1号,它长约5公里,该区的五个最大的工业矿体关门山矿段21、22号,小西沟矿段1号、2号、7号等矿体均受此断裂控制,因此认为该区为典型的裂控矿床,故应沿断裂找矿。

我们采用的找矿方法是构造化探,即对矿区内所有断裂采样。在同一断裂上,每隔300米左右布置一剖面,该剖面近似垂直于断裂,剖面长度略大于断裂破碎带宽度,每条剖面所采样品数不等,以能反映成矿元素沿断裂带变化特征为目的。在该区共采88条(包括已知含矿地段的9条。各剖面位置见图1)计1414个原生晕样,分析了9个项目(Pb、Zn、Cd、Ag、Hg、Sb、As、Bi、Cu),我们从中选了63条剖面,各剖面编号、名称、样品数及属性见下表1):

对上述63条剖面的958个样品,用离差平方和法分群,其结果见谱系图(见图2):

由谱系图可以看出,63条剖面可以分成两大类,第一大类是(参照表1):

8号—NE ₁ P ₁	31号—NW ₁₀ P ₅
9号—NE ₁ P ₂	32号—NW ₁₂ P ₁
11号—NE ₁ P ₄	59号—CP ₁
15号—NE ₁ P ₈	60号—CP ₂
23号—NE ₅₀ P ₁	

其中的NE₁P₁、NE₁P₂、NE₁P₄、NE₅₀P₁、CP₁、CP₂为典型的已知含矿区剖面,故ZE₁P₈、NW₁₀P₅、NW₁₂P₁可定为一级靶区。

再看第二大类,它们还可分为两小类。第一小类是:

1号—EW ₃ P ₁	37号—NW ₁₈ P ₂
7号—NE ₁₀ P ₁	46号—NW ₂ P ₃
24号—NE ₅₀ P ₂	49号—NW ₃ P ₁
29号—NW ₁₀ P ₃	54号—NW ₄ P ₁
34号—NW ₁₇ P ₂	57号—NW ₉ P ₂
35号—NW ₁₇ P ₃	

其中EW₃P₁、NE₅₀P₂为已知含矿区剖面,与它们相似的可定为二级靶区。其余所有剖面,与模型区相似程度低,暂不能定为靶区。

为了验证该方法的分类结果是否合理,我们从地质上来分析上述各级靶区的特点:

NE_1P_8 :它位于本区最重要的已知控矿断裂 NE_1 上,岩性为典型的条带状白云岩,南有辉绿岩屏蔽。从原生晕来说,主要成矿成晕元素 Pb、Zn、Cd、Ag、Hg 异常高,Pb、Zn 最大值均在 1000ppm 以上,异常范围大,故为较理想的一级靶区。

$NW_{10}P_3$ 、 $NW_{10}P_5$:该范围为 NW_{10} 与 NE_2 交汇处,北有辉绿岩屏蔽,异常范围大,主要成矿成晕元素含量高,可定为一级靶区。

$NW_{12}P_1$:位于关门山矿段东北部,为条带状白云岩分布区,南有辉绿岩屏蔽。就化探异常来说,主成矿元素及伴生元素均高,是一个较好的一级靶区。

NW_3P_1 :位于南山,岩性为条带白云岩,处于两组断裂交汇处,有 Pb、Zn、Ag 异常,但强度不大,可作为二类靶区。

$NW_{17}P_2$ 、 $NW_{17}P_3$ 、 $NW_{18}P_2$:位于崔北沟南山,为多级断裂交汇处,有弱的铅锌异常,可作为二类靶区。

NW_2P_3 :位于老虎洞,为两组断裂交汇部位,主要成矿、成晕元素均有异常,定为二类靶区。

$NE_{10}P_1$ 、 NW_9P_2 :为成矿条件差的二类靶区。

由上面的分析可知,用样品集合分类所得的结果与地质上的基本吻合。

最后要说明的是:样品集合分类不仅可以用于对异常进行评价,还可以用于对矿体、岩体等地质体进行分类。

这篇文章在写作过程中,得到徐锡华工程师、朱作山工程师的热情指导和帮助,在此表示感谢。

参考文献

[1]煤炭研究所,西安矿院数学教研室,数学地质基础与方法,煤炭工业出版社,1979

APPLICATION OF GROUP CLASSIFICATION TO ASSESEMENT OF GEOCHEMICAL ANOMALIES

Hu Kun

Abstract

This paper introduces a new sample classification--group classification. It's advantage is that each point in plot represents a geological body rather than a sample point. It has many math models. Here is used the model of the sum of dispersion square. Application of this method to Guanmenshan Pb-Zn Mine obtained the same results as geological analysis.