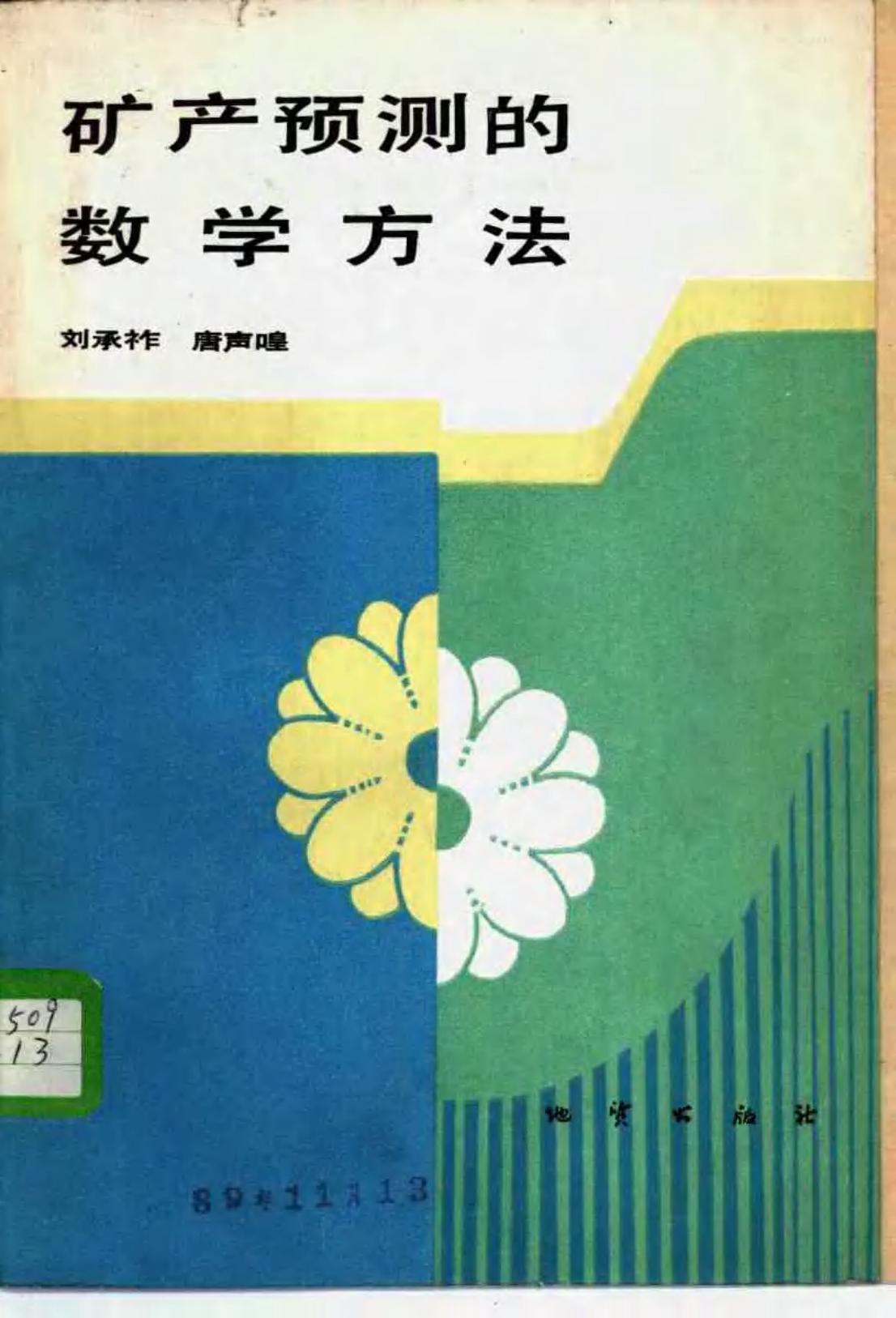
矿产预测的数学方法

刘承祚　唐声喤



地质出版社

矿产预测的数学方法

刘承祚　唐声晅

地质出版社

内 容 提 要

本书主要介绍一些较新的预测矿产的数学方法，并附有很多实例。书中收入了大量的国外预测矿产的实例，同时也介绍了作者的一些工作经验。这些对矿产资源预测工作都具有一定意义。

本书可供一般地质人员和地质院校师生参考，也可作为专业数学地质工作者的参考书。

矿产预测的数学方法

刘承祚　唐声鸣

*

责任编辑：杨友爱

地质出版社出版发行

（北京西四）

地质出版社印刷厂印刷

（北京海淀区学院路29号）

新华书店总店科技发行所经销

*

开本：850×1168 1/32 印张：6.0625 字数：159,000

1989年2月北京第一版·1989年2月北京第一次印刷

印数：1—1430册　　国内定价：2.35元

ISBN 7-116-00405-X/P·350

序　言

应用数学的方法预测矿产是数学地质的一个新分支。这种工作在国外大约出现于50年代末期和60年代初期。随着地质研究定量化水平的提高和数学地质的发展，矿产地质学家在实践过程中对预测矿产的数学方法愈来愈重视，研究这类问题的人也日益增多，近年来已进入高潮。在最近几届国际地质大会的数学地质分组会上，召开过一系列矿产资源定量评价预测的专题讨论会，宣读了许多很有价值的学术论文。国内的矿产预测工作开始于70年代中期。在第一届、第二届和第三届全国数学地质学术会议的论文中，涉及矿产预测方面的占有很大比重，可见国内数学地质工作者对此问题的重视。我国对这方面的研究工作，目前正朝着一定的广度和深度发展。

本书综述了大量有关预测矿产的数学方法，并重点介绍了文献〔29〕、〔31〕和〔32〕中提及的内容，同时也介绍了本书作者从事这方面专题研究所取得的一些成果和经验，供国内同行参考。

本书的一、四、五、六、七章由刘承祚执笔；第二、三章由唐声嗤执笔；柴俊杰参加了第五、六章的部分工作；陶起、英定文参加了第六章的部分工作。刘承祚完成了全书的汇编整理工作。全书完稿后，承蒙核工业部研究员姚筱煌审阅并提了许多宝贵意见，谨在此表示感谢。

限于作者水平，对本书中的错误和不足之处，敬请读者批评指正。

作　者

目　录

第一章 绪 论

随着我国国民经济的日益发展，国家对各种矿产资源的需求日益增大。因此，地质工作者必须深入开展矿产预测研究，探索预测矿产的新方法（包括数学方法）。

§1. 矿产预测的一般原理

为了成功地预测矿产，必须十分重视地质基础工作，在此基础上开创性地进行成矿理论研究，并根据新的理论对矿产预测予以指导，同时重视引进新技术和新方法。

1. 新成矿理论对矿产预测的指导意义

新成矿理论的出现曾多次使找矿方向发生重大改变，导致了大量新矿床的发现：层控和层状矿床理论的出现，促使发现了许多层控和层状新矿床；地层-岩性圈闭理论指导了石油矿产的预测；洋流上升理论帮助人们找到了大型磷矿床；火山成矿作用理论对含铜黄铁矿型矿床和黄铁矿型多金属矿床的预测起了重大指导作用等。因此在矿床预测工作中应充分应用新的理论。

2. 控矿地质因素研究是矿产预测的重要基础

控制矿产分布的各种地质因素的研究是在一个地区开展矿产预测的前期工作。控矿地质因素主要包括构造控矿因素、岩浆岩控矿因素、变质控矿因素和表生控矿因素等，这已为地质工作者所熟知，不拟多述。

3. 新技术和新方法的综合应用是提高矿产预测的重要途径

目前，未被发现的地表露头矿已愈来愈少，用传统找矿方法找到的矿床也越来越少。在世界绝大多数地区，找矿工作的重点已转移到寻找隐伏矿体和被一般找矿方法可能遗漏的矿床方面。

在这种新形势下，为了成功地找到矿床，必须应用多种新技术和新方法，收集尽可能多的直接的和间接的找矿信息（标志），对其进行综合研究，以指导预测。这些新技术和新方法主要包括：物探、化探、航空地质、遥感地质、数学地质以及更高精度的测试分析法等。这些新技术应根据所预测的矿床类型，有针对性地加以综合应用。

数学方法是矿产预测的重要新手段之一。数学方法能够充分发掘找矿信息并对其进行综合分析，充分发挥这些信息的作用。因此，与传统矿产预测相比，数学地质方法有一定的特点和优势。

§2. 应用数学方法预测矿产的发展概况

应用数学方法预测矿产大致经历了三个发展阶段。

1. 初始阶段（1957—1964年）

在此阶段，地质学家开始应用数学方法预测矿产工作。如何将地质因素与数学模型有机地结合起来，还处于探索阶段。所建立的数学模型大都较简单，预测矿产的效果不理想。

在这一阶段比较有代表性的模型是由阿雷斯（M.Allais）所建立的。1957年，他对美国西部的含矿盆地进行矿产预测时，把预测区划分为许多单元，每个单元面积为10km²。他发现每个单元中的矿床数目服从泊松分布：

$$P(x)=\frac{e^{-\lambda}\lambda^{x}}{x!}$$

式中：$P(x)$为单元中发现x个矿床的概率，λ为每个单元中发现矿床的数学期望。

阿雷斯应用上述结论对阿尔及利亚的撒哈拉地区进行了预测。当时根据已知资料，认为这是一个与他研究的美国西部盆地具有相似地质条件的地区。预测前未对预测区进行踏勘。预测后认为在预测区有可能找到20个矿床，平均每5km²面积内能找到

一个矿床，并根据矿床经济价值的概率分布（一般假设为对数正态分布）和勘探投资额进行了计算。其结论是：用200亿旧法朗勘探投资能找到净值为500亿法朗的矿产的概率为0.65。

阿雷斯模型是最早应用于矿产预测的数学模型。这个模型有下列特点：(1)比较简单，仅把矿床数目分布外推到未知区进行预测；(2)未考虑地质因素和矿产之间的关系；(3)所利用的地质信息很少，结果不准确。

除此之外，还有一些其他的简单数学模型，不一一列举。

2．趋于完善阶段（1965—1974年）

在此阶段，应用数学方法预测矿产工作得到了进一步发展，并取得了一定实效。所建立的数学模型日趋完善，矿产预测的实际效果也比较理想。电子计算机技术的引进，为矿产预测工作提供了重要的研究手段。

在此阶段值得一提的是哈里斯（D.P.Harris）所建立的地质数学模型，其主要内容包括：

（1）选择勘探程度较高的亚利桑那州和新墨西哥州的一部分作为已知区进行研究。

（2）将已知区划分为240个单元，每个单元面积为32km²。

（3）从研究区的地质图上选取了24个地质变量。这些地质变量可划分为下列四大类：

A．岩石类型和时代　单元中第三纪以前沉积岩所占百分数和前寒武纪侵入岩所占百分数等；

B．断裂　长度为0—8英里[1]的高角度断裂数目和高角度断层的交叉数目等；

C．构造型式　长度为0—8英里的背斜数目和长度大于8英里的背斜数目等；

D．侵入活动的时代和接触关系　拉拉米和内华达侵入岩同沉积岩接触带的长度，前寒武纪侵入岩同寒武纪变质岩接触带的

[1] 1mile（英里）＝1609.344m

长度等。

（4）根据过去的矿山开采资料，计算每个单元的矿产资源价值，大于100万美元的划为第一组，小于100万美元的划为第二组。

（5）建立两组判别函数　确定26个地质变量和矿产资源组（第一组和第二组）之间的关系。当归入第一组的概率大于0.2时，即可认为该单元是有利于进一步勘探的地区。

（6）从地质条件类似的犹他州部分地区，选出了144个单元进行验证。验证结果预测19个单元为有矿单元。将这19个单元与犹他州已知矿床比较，结果如下：

矿床级别（按累计总产值计算）	预测矿床数	实际矿床数
>1亿美元	5	5
1000万至1亿美元	2	4
100万至1000万美元	3	8
<100万美元	9	0

哈里斯模型的主要特点是：(1)它是一个考虑了地质因素的比较完善的模型；(2)矿产预测的实际效果较好。

在综合应用阿雷斯模型和哈里斯模型的基础上，阿格特伯格（F.P.Agterberg）根据每个小单元的地质特征用回归分析的方法估计小单元的矿床发现概率，然后利用这些概率值建立一个大单元（由若干小单元组成）内矿床数目的泊松分布模型，进行矿产预测。

1973年哈里斯提出了主观概率模型。他在德尔菲（Delphi）方法的基础上设计了一种应用数学方法和电子计算机整理、归纳地质人员在矿产预测方面集体意见的方法，进行矿产预测。

3．蓬勃发展阶段（1975年至现在）

在这个阶段，数学方法已被应用于各种矿种和各种比例尺的矿产预测，应用的深度和广度也不断提高。用于矿产预测的数学模型更趋成熟，对实际工作取得了较好的效果。

在这一阶段中，有下列重要事件值得特别提出：1975—1980

国际地质相关计划（IGCP）第4组设置了第98号专题“资源研究中计算机应用的标准”，总结推广了六种矿产预测方法（单位区域价值估计法、丰度法、体积法、德尔菲法、矿床模型法及综合法）。这一专题的设置对矿产资源数据的搜集、处理和应用标准化起了很大推动作用。

在1976年召开的第25届国际地质大会上，国际数学地质协会召集了“定量勘探策略”专题讨论会，交流了应用数学方法预测矿产的经验。在26届和27届国际地质大会中，对应用数学方法预测矿产专题进行了讨论。

苏联学者康斯坦丁诺夫（Р.М. Константинов）1977年应用逻辑信息法预测矿床的可能规模，1980年布加耶茨(А.Н.Бугаец）应用模糊数学理论进行了矿床预测。

在我国，应用数学方法预测矿产的工作开始于1975年。首先在宁芜盆地中段和北段应用多元统计方法和数理统计方法开展了铁矿预测，随后又将多元统计方法用于其他地区的铁矿预测。1977年，统计对策方法被用来在安徽省庐枞盆地北段预测铁矿床。以后各个工业系统的地质单位都开展了应用数学方法预测矿产的工作，并取得了大量成果。

§3. 应用数学方法预测矿产的基本工作环节

1. 资料收集和数据准备

目前应用数学方法预测矿产的工作都是在现有地质资料基础上进行的。这些资料包括各种比例尺的地质类图件和文字资料、物化探资料、遥感资料、各种类型的矿产资料（包括区域矿产统计资料）以及地质人员的经验资料等。各类资料中有一部分是定量的，可以直接使用，另外一部分是定性的，首先需将其转换成定量的、数字的形式。形成数据后按照统一化和规格化的数据格式，对数据进行收集、整理和登录。数据经过加工、分析和处理，将其中包含的大量有用信息提取出来，用于矿产预测。

2．预测单元的划分和研究变量的选取

预测矿产使用的数学方法绝大部分都要求将研究区域划分成若干单元，根据矿种和预测比例尺确定单元的适宜面积。单元面积大小选择合适与否对矿产预测效果会产生重要影响。单元分为两大类：已知单元和未知单元。已知单元再进一步划分为有矿已知单元和无矿已知单元两个亚类。此外，还要根据地质条件和预测矿种选择适当的研究变量。变量可以是地质的、地球物理的，也可以是其他与成矿有关的因素。变量可以是定量的，也可以是半定量的或定性的。

3．由已知区到未知区的推断

在绝大多数情况下，应用数学方法预测矿产都是从研究地区的已知区外推到未知区，这时，要求已知区和未知区的基本地质条件相似。首先根据已知区的地质条件归纳出地质概念模型，进而建立数学模型，将所建立的数学模型外延应用到未知区，以达到对未知区进行矿产预测的目的。

4．数学模型的建立

根据地质模型以及选定的变量和矿产预测要求，建立恰当的数学模型，这是预测工作的关键性步骤。数学模型是否符合已知区和未知区的地质条件以及客观成矿规律，又是矿产预测成败的关键。成功的数学模型不一定非常复杂，但要反映出控制矿产分布的最重要的地质因素和指示矿体存在的最重要的那些信息。

5．矿产预测结果的地质解释和野外检验

对用数学方法预测矿产所取得的结果应做出合理解释并进行野外验证，看其结果是否正确。验证的方法很多，如在圈出的预测靶区内进行加密化探，补充物探以及进行钻探等。

第二章　直观推断法[1]与自动分类法在矿产预测中的应用

§1. 直观推断法

§1.1 几个概念　直观推断法（启发法）属于类比法的一种。应用该方法时，需要各类（二类或多类）已知标准对象若干个。这种方法只适用于处理定性资料，其判别规则并非根据某些准则（如费歇准则、贝叶斯准则）进行数学推导得到的。而是根据符合思维逻辑的原则，在已知标准对象中寻找一些有重要标志的各种组合（如终止行异列集票数法）或有某三种标志的特定组合（如地壳-3法），然后在未知对象中，寻找与上述有相同标志的组合数。最后，把未知对象判归为所述组合数符合最多的那个已知对象。这种方法由于包含了地质工作者根据找矿标志进行预测的思想，而又不受一组或为数不多的几组找矿标志的局限，所以其预测效果是明显的。

应用直观推断法进行矿产预测时，需要用终止行异列集。为此，引进如下几个概念。

适用表　设有m个对象$x_1, x_2, \cdots, x_m$，每个对象用n个二进位数来描述：即$x_i=(\alpha_{i1}, \alpha_{i2}, \cdots, \alpha_{in})$，$\alpha_{ik}=0, 1$；$i=1, 2, \cdots, m$；$k=1, 2, \cdots, n$，则对象$x_1, x_2, \cdots, x_m$可用表

$$T=\begin{pmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2n} \\ \alpha_{m1} & \alpha_{m2} & \cdots & \alpha_{mn} \end{pmatrix}$$

来表示。若在表T中，每一列至少有一个1和一个0，且没有二行是完全相同的，这样的表称为适用表。

---

[1] Эвристический

行异列集　在适用表中，选任意几列组成各种可能的列组合，把那些没有二行是相同的组合称为行异列集。

终止行异列集　在行异列集中，当去掉任意一列后，就出现了相同的二行（即不再是行异列集），把这种行异列集称为终止行异列集。

多表行异列集　有多张（二张以上）列数相同的适用表，每一张表代表一类对象。任选几列组成各种可能的列组合，把那些在不同类之间没有二行是相同（在同类间二行允许相同）的组合称为多表行异列集。

多表终止行异列集　在多表行异列集中，当去掉任意一列后，就在不同类中出现了相同的二行(即不再是多表行异列集)，把这种多表行异列集称为多表终止行异列集。

§1.2　终止行异列集票数法（简称《ΓT》法）　设适用表T是由p个不同类型的标准对象（以行表示）及描述它们的n个特征（次列表示）所组成。由该表求出的M个终止行异列集分别为T_1，T_2，…T_m。若表中每行用$(a_{i1}, a_{i2}, \cdots, a_{in})$表示，其中$a_{ij}=0, 1$，$i=1, 2, \cdots, p$；$j=1, 2, \cdots, n$。现有未知对象$Q$个，分别以$(b_{k1}, b_{k2}, \cdots, b_{kn})$表示，其中$b_{kj}=0, 1$，$k=1, 2, \cdots, Q$；$j=1, 2, \cdots, n$。如欲对某个未知对象$U_k$判断其归属，则可根据诸终止行异列集合出的总票数$g_1^{(m)}$，$g_2^{(m)}$，…，$g_p^{(m)}$多少来决定。票数的计数法如下：对任意一个终止行异列集而言，若待判的一个对象在这个终止行异列集所包含的列号的取值（1或0）与某个标准对象在这些列号的取值完全相同时，则这个标准对象得一票，依次把全部终止行异列集照此做完为止，然后统计各标准对象按全部终止行异列集共得的票数。待判的对象应归属于得票最多的那个标准对象。用公式表示如下：

$$g_i^{(0)}=0;\quad g_i^{(r)}=\begin{cases} g_i^{(r-1)}+1 & \text{对于第 } r \text{ 个终止行异列集，其第 } i \text{ 行诸特征的取值与待判对象的相应特征的取值完全相同} \\ g_i^{(r-1)} & \text{除上述外的其它情形} \end{cases}$$

$$i=1,\ 2,\ \cdots,\ p;\ r=1,\ 2,\ \cdots,\ M$$

若$g_{i_0}^{(M)}=\max\{g_i^{(M)}\}$，则待判对象归入第$i_0$类。

§1.3 多表终止行异列集票数法（简称《ГТР》法） 当标准对象有二大类且每类包含若干对象时，此时便有二张列数相同的适用表T_1,T_2。若由该二表求得的多表终止行异列集数为M'则计算多表终止行异列集票数的算法和上述类似，而待判对象应归入得票最多的那一张表所代表的类别。用公式表示如下：

$$g_i^{(0)}=0;\ g_i^{(r)}=\begin{cases} g_i^{(r-1)}+1 & \text{对于第 } i \text{ 个终止行异列集而言，若第 } i \text{ 张表中的任一行在这些列（特征）的取值与未知对象相应特征的取值完全相同} \\ g_i^{(r-1)} & \text{除上述外的其它情形} \end{cases}$$

$$i=1,\ 2;\ r=1,\ 2,\ \cdots,\ M$$

若$g_1^{(M)}>g_2^{(M)}$则待判对象归入第一类，反之则归入第二类。

§1.4《地壳-3》算法（简称《k》法） 直接推断法中也包括《地壳-3》算法，简称《k》法，该方法原理如下：

设以二个集合$x_1=\{x_1^{(1)},\ x_2^{(1)},\cdots,x_N^{(1)}\}$和$x_2=\{x_1^{(2)},\ x_2^{(2)},\ \cdots,\ x_{N_2}^{(2)}\}$表示二类对象。每个对象用$n$个指标$x_1$，$x_2$，$\cdots$，$x_n$来描述。若在第一类中$p$个对象的相应指标（$x_p$，$x_q$，$x_r$）的值为（$\alpha_p$，$\alpha_q$，$\alpha_r$），而在第二类中却没有一个对象之相应指标具有此值，则指标三元组（x_p，x_q，x_r）称为充分指标。

设在1类中，有p_1个对象的某三个指标均取值（α_{p_1}，α_{q_1}，α_{r_1}），而有p_2个对象的另三个指标均取值（α_{p_2}，α_{q_2}，α_{r_2}）。若$p_1>p_2$，且在p_1个对象中包括了p_2个对象，则可认为三元组（α_{p_1}，α_{q_1}，α_{r_1}）所对应的充分指标要比（α_{p_2}，α_{q_2}，α_{r_2}）所对应的强些，并称其为强充分指标。应用充分指标对未知对象进行判别分类时，需要预先去除强充分指标。

求第一类对象中的充分指标数直至其值达到预先给定的数l_1为止。如果充分指标数达不到l_1，则可降低阈值p（对象个数）。

然后按新阈值继续求该类对象中的充分指标数。这样的计算过程可延续到p等于1为止。用同样的方法求第二类对象的充分指标数l_2。

在充分指标数l_1及l_2的坐标系中，判别未知对象$Y=(y_1, y_2, \cdots, y_n)$归属的方法如下：若由该对象求得的与$l_1$及$l_2$个充分指标数中相符合的个数为$r_1$及$r_2$，且$r_1>r_2$，则未知对象判归为1类，当$r_1<r_2$，则判归为2类。

§2. 自动分类法

自动分类法和其它的数学预测方法不同，它适用于仅有一类标准对象或完全没有标准对象时的情况。

应用自动分类法进行预测的问题可表述如下：根据描述各对象的诸指标x（向量），计算各对象之间的相似系数（或距离）。在给定组（类）内的相似系数I_1和组间的相似系数I_2后，为了把对象进行分类，必须使I_1达到极小而使I_2达到极大。这样就能把诸对象合并成预先确定的分类数。

自动分类法所包括的算法不下数十种（可参阅聚类分析专著），但在矿产预测中最常用的有《谱》、《合并》和《二二合并》等三种。下面仅对这三种算法的原理作一简述。

§2.1《谱》算法 《谱》算法的计算步骤如下：

第一步：根据归纳法，采用二个序列$L_N=(\tilde{x}_1, \tilde{x}_2, \cdots, \tilde{x}_N)$和$Q_N=(\tilde{d}_1, \tilde{d}_2, \cdots, \tilde{d}_N)$与原始资料向量集$M_0=(x_1, x_2, \cdots, x_N)$相对应。由归纳法：

首先，设$\tilde{x}_1=x_1$，$\tilde{d}_1=1$，从M_0中去除x_1得到的集合设为M_1。

当计算工作进行到第$m+1$步时，从原始的M_0中已选出m个向量，并建立了二个序列：$L_m=(\tilde{x}_1, \tilde{x}_2, \cdots, \tilde{x}_m)$和$Q_m=(\tilde{d}_1, \tilde{d}_2, \cdots, \tilde{d}_m)$，而其余的向量合并在集合$M_m$中，故在进行第$m+1$步时，可建立序列：

$$L_{m+1}=(\tilde{x}_1,\ \tilde{x}_2,\ \cdots,\ \tilde{x}_m,\ \tilde{x}_{m+1})$$

$$Q_{m+1}=(\tilde{d}_1,\ \tilde{d}_2,\ \cdots,\ \tilde{d}_m,\ \tilde{d}_{m+1})$$

其中

$$\tilde{x}_{m+1}=\{x_{j_0};\ \max\ d(L_m,\ x_j)\}$$

$$\tilde{d}_{m+1}=d(L_m,\ \tilde{x}_{m+1})$$

距离 d 的计算可采用类平均法、最短距离法或最长距离法中任何一种方法。

因此，第一步的计算实质是：首先，按某种原则（任意选定或用某个超球面中心）确定向量 $\tilde{x}_1$，并从其全部 $N-1$ 个向量中挑选与 $\tilde{x}_1$ 最接近的 $\tilde{x}_2$，然后，再从 $N-2$ 个向量中挑选与 $\tilde{x}_1$ 和 $\tilde{x}_2$ 最接近的 $\tilde{x}_3$。在计算过程中，要使 $\tilde{x}_j$ 与 $\tilde{d}_j$ 相适应。

第二步：对序列 $\tilde{x}_1$，$\tilde{x}_2$，…，$\tilde{x}_N$ 进行初始分类，并分成 x_1，x_2，…，x_{KH} 等几类。为了进行初始分类，需要把序列 $\{d_j\}$ 变换成序列 $\Delta=\{\Delta_j\}$，

$$\Delta_j=(\tilde{d}_j-\tilde{d}_{j+1})/\tilde{d}_j \quad j=1,\ 2,\ \cdots,\ N-1$$

从序列 Δ 中挑选 $a-1$ 个最大值 $\Delta_j(a=k_H)$ 并将这些值由小到大排列。这样得到的序列以

$$\Delta_{1,p}=\{\Delta_{1,p}^{(k)}\} \quad k=1,\ 2,\ \cdots,\ a-1$$

表示。

然后按下列步骤把序列 $\{x_j\}$ 分成 d 类。

首先，设 $\tilde{x}_1\in x_1$。

如 $\tilde{x}_{j-1}\in x_1$，当 $\Delta_j<\Delta_{1,p}^{(1)}$ 时，则 $\tilde{x}_j\in x_1$；若 $\Delta_j\geqslant\Delta_{1,p}^{(1)}$，则 $x'_j\in x_2$。此时，分在 x_1 类的分类工作结束（该类中进入了 $\tilde{x}_1$，$\tilde{x}_2$，…，$\tilde{x}_{j-1}$）。

用同样的方法，根据 $\Delta_{1,p}^{(1)}$，$\Delta_{1,p}^{(2)}$，…，$\Delta_{1,p}^{(a-1)}$ 的值，就可得到其它的类别。

§2.2 《合并》算法 该方法的原理如下：设有 N 个点 x_1，x_2，…，x_N（每个点代表一个对象或一个样本），现要求把这 N 个点合并成预先给定的分类数 r_0。今假设并类工作进行到第 l 步，即把 $\{x_i\}$ 合并为 A_1，…，A_{v_l} 等 v_l 类，在作第 l 步计算时，要求把 A_i 和 A_j 二类合并为一类。进行合并的原则是要使该二类间