

<http://www.geojournals.cn/georev/ch/index.aspx>

矿产预测中成矿信息的提取和强化

浦志伟 朱裕生

(中国地质科学院成矿远景区划室, 北京)

成矿信息的提取是矿产预测的重要手段, 其方法的有效性直接影响到预测成果的可靠性。在成矿规律研究的基础上, 有意识地干预模型的构成, 突出与成矿有关的信息; 抑制某些成矿意义不明显和属干扰的信息, 有目的地使模型向反映成矿信息的方向逼近, 提高模型与矿床实际赋存地质环境的吻合程度, 均是成矿信息提取和强化的内涵。本文从研究实践中总结了先验约束模型和非先验约束模型的强化方法, 在新疆阿勒泰地区的地质-找矿工作中已取得实效。



关键词 成矿预测 成矿信息的提取 成矿信息的强化与转换

1 概述

在一定条件下, 矿与非矿之间存在某种程度的差异, 这种差异就向人们提供了认识矿床的信息。能够指示或识别某种矿床成矿条件或赋存方式的资料、数据和综合参数都可称成矿(含矿)信息^[1]。随着矿产预测工作的逐步展开, 成矿信息的概念也越来越为地质工作者所接受。然而, 根据成矿系列理论, 在一定地质构造单元内和一定地质发展阶段中, 与一定地质作用有关的产于不同构造部位的不同类型的矿床(种)组在时、空分布上有密切联系^[2]。属于同一系列或亚系列的矿床必定会在成矿信息方面表现出某种程度的相似性^[3]。另外由于地质科学发展的阶段性和对地质规律认识的不完善性, 人们尚不能把握住成矿的充分条件。通过地质研究, 我们总可以归纳出一个(类)矿床比较显著的赋矿或控矿标志, 这些标志似乎可以作为寻找这类矿床的必要条件。但有些成矿信息是比较隐蔽的, 不易被识别, 需要数据处理、综合分析或人为强化才能显示。因此, 就目前的研究水平而言, 矿产预测只能是对成矿有利度的预测, 而不是直接对矿床或矿体的预测。

矿产预测的方法多种多样^[1], 但其理论前提仍然是类比。简单类比只是将某些直观的成矿标志直接进行对比; 而综合类比则经过对成矿标志的综合研究, 抽象成为模型, 然后再将研究对象与模型进行对比, 继而对研究对象作出某种判断。而数学模型的引入为综合类比的技术方法及矿产预测的发展奠定了基础。

最大限度地浓缩和提取成矿信息是矿产预测的一项重要任务。矿产预测的过程实际上是在探索成矿规律的基础上, 通过模型将各类地质矿产信息进行转换、合成(综合)及筛选、对比^[4], 对潜在矿床作出判断, 达到预测目的。一种有效的成矿预测方法, 不仅能比较客观地反映矿床的成矿环境和成因机理, 并对有关的控矿标志进行组合和赋权, 而且还能突出一些比较隐蔽的、有地质意义的成矿标志信息。在这里, 数学模型只相当于变换器, 经过它的作用, 可将具有错综复

杂相关关系的地质信息转化为有一定成矿意义的成矿信息。

矿产预测使用的方法是否有效,主要取决于其能否反映成矿模式的地质意义和最大限度地突出成矿信息。但在应用过程中,很难使地质模型和数学模型两者有机结合。其主要原因是:1.地质(矿床)模型的复杂性,矿床的形成往往需要经过多次成矿地质作用过程、受多种复杂因素的制约,数学模型不可能完全描述漫长地质年代里发生元素富集或分散的全过程;2.缺乏统一的类比前提,例如已知矿床分布模型区与预测区的地质工作程度存在明显差异,这种差异不是矿产预测过程中可加以弥补的;3.数学模型的严密性,数学模型应用条件是严密的,如果不符合模型的适用前提,其预测结果可能是十分荒谬的。目前的地质研究尚处于定量化程度较低的阶段,相应的地质数据很难满足或只能勉强满足模型的边界条件。因此,其预测只具有一定的可信度,据此现状,本文提出了成矿信息强化的方法,视为提高成矿预测成果可靠性的途径。

2 成矿信息强化的途径和模型的选择

所谓成矿信息强化是人们根据对成矿地质规律的认识,有意识地干预模型的构成,突出其中的一些与成矿密切相关的标志(组合)信息,抑制某些成矿意义不明显或干扰较强的信息。当然,这种人为干预是在一定的预测准则前提下,有目的地使模型向反映成矿信息的方向逼近,提高模型与矿床实际赋存条件的吻合程度。

数学模型可通过数量或图形再现客观规律,但并非完全真实地反映,两者间的差异往往是衡量模型适用程度的标尺。这里所说的强化实际上就是强化模拟,以此缩小模型与模拟对象(矿床实体)之间的差异。当前用于矿产预测的数学模型、地质模型是多种多样,但是总体上看,根据模型是否受先验条件约束可分为先验约束模型和非先验约束模型。

先验约束模型是指模拟对象的某些属性已经确定,并且以模拟这些属性为主要目标的模型。如回归模型用因变量的特征对模型进行约束;判别模型则先将研究对象合成不同的类或组,使信息的属性具有明显的倾向性;逻辑信息模型也有类似的特性。而非先验约束模型则没有这种先决条件的约束,它们通过变量相互关系的表征,反映出不同层次的单一或组合信息,其结果往往需要结合地质实际进行解释后才能显示出它的成矿地质意义,如因子分析模型、特征分析模型、聚类分析模型等。如果说前一类模型能直接反映某些方面的成矿信息,那么后一类模型则是对成矿信息的间接反映。当然这二类模型的界限不是绝对的,只要对这些模型稍加改造或修正便可使其成为另一类模型。

基于这二类模型在先验约束条件方面的差异,对它进行信息强化的目的和方法手段也不尽相同。对于第一类模型,强化的目的就是最大限度地提高模型的拟合程度,即被模拟的特征参数用地质模型所提供的各种变量来描述,使两者的差异达到最小,其差异程度常用一些特征的数学参数来衡量,如回归模型的复相关系数、趋势面模型的拟合度等;对于第二类模型,要求在充分解释构成模型各类变量组合及其相应成矿信息意义的基础上,研究已知模型区的成矿信息特征及与已知矿床(体)关系的密切程度,并以此为出发点,将全部预测对象的组合变量标志向有利于模型特征方向转换(有利定向转换),然后将成矿信息浓缩,使其成为反映模型成矿意义的有利信息或预测标志。

3 先验约束模型

就矿产预测来说,先验约束模型是在已知区内矿床的规模及其相应属性已知的情况下建立的

预测模型。从以下实例中可以探索这类模型的实际应用和改进的途径。

阳春盆地是在印支复式向斜基础上发育而成的小型断陷盆地, 并受北东向深大断裂所控制。区内各时代地层出露齐全, 褶皱断裂极为发育, 岩浆岩不仅分布广泛, 而且岩性复杂, 以铜和锡、钨多金属为主。在研究成矿规律的基础上, 确定17个地质标志与区内的夕卡岩型、高温热液型铜(钨)矿床密切相关。其中, 只有5个化探元素是半定量光谱分析数据, 而其它的地质标志只能表现为存在与否的定性变量, 数据的结构和性质有较大差异。另外, 在所确定的27个模型单元中, 探明储量相差悬殊, 除一个单元中已知矿床的探明储量达到大型规模以外, 其余均为矿点(或矿化点)。在此前提下, 建立因变量(探明储量)与自变量(地质条件)之间的回归模型或数量化理论I模型, 很难取得实效, 如实际计算的复相关系数仅为0.5862。由于应用的数学模型(此处是多元回归模型)缺乏应有的稳定性, 用它预测时, 出现一些不可思议的“奇异值”。如果采用稳健回归方法, 必定使模型中大型矿床的作用受到削弱, 将影响到全区预测储量的估算。

从上述实际问题出发, 我们提出了成矿信息强化QMBD(Quantative Method Basted On Dependent Variable的缩写, 原意为基于因变量信息的定量方法)模型, 用它实现模型的定向转换。

QMBD模型侧重点是对定性自变量的修正, 改变原来自变量与因变量处于不同信息等级的状态。建立QMBD模型的基本解释是:

1. 在现有的研究水平下, 模型区的探明储量与反映该区成矿作用的地质标志有关, 且只与这些标志有关;

2. 某单元的预测储量与其M个成矿地质条件之间呈如下线性关系:

$$Y_i = \sum a_j x_{ij} + s_i \quad (1)$$

其中: Y_i 为第*i*单元的探明储量, $i=1, 2, 3, \dots, n$; a_j 为第*j*个变量权, 是待定系数, $j=1, 2, 3, \dots, m$; x_{ij} 为定性地质变量经转换后的定量潜在资源信息, 是QMBD的内涵; s_i 为随机误差。

(1)式与回归模型不同之处是未出现常数项。

(1)式的意义就是把*i*单元的预测储量(此处与潜在资源量意义同等)看作是该单元各变量对预测储量的贡献—— $a_j x_{ij}$ 及 s_i 的总和, 用*n*个模型的已知探明储量和*m*($< n$)个地质标志(定性变量)来估计 x_{ij} 的权 a_j 。其关键问题是要确定 x_{ij} 的意义及表达方式。

为了简便起见, 我们假设原始数据所确定的定性变量为二态变量, 即若单元内存在第*j*个地质标志, $\delta_{ij}=1$; 否则 $\delta_{ij}=0$ 。按上述假设, 我们通过已知的探明储量的信息使同变量中的相同取值得到区分, 也就是强化加权。设函数:

$$X_{ij} = f(\delta_{ij}, y_i) = \frac{y_i \delta_{ij}}{\sum y_k \delta_{kj}} \sum y_k \quad (2)$$

其中: $\sum y_k$ 为模型区的已知探明储量。

(2)式前面的分式是出现第*j*个地质标志单元的探明储量在所有含该地质标志单元的探明储量总和中所占的比重, 它表明了各单元 δ_{ij} 的相对重要性。若 $\delta_{ij}=0$, 则 x_{ij} 亦为0; 但若 $\delta_{ij}=1$, 则对同一变量来说, 为同单元中取值相同的“1”却会有不同的反映, 可以看出 y_i 越大, 则 x_{ij} 也就越大, 这就出现了这个“1”比那个“1”更重要的状态。从(2)式可知, 若单个变量恒为“1”, 则它与因变量之间呈完全相关关系, 若有“0”出现, 则两者之间就不可能完全相关, 但其相关程度要比未作这种变换时高得多, 这是极明显的。进而分析 $a_j x_{ij}$ 的意义将出现如下形式:

由(1), (2)式可知:

$$a_i x_{ij} = a_i \frac{Y_i \delta_{ij}}{\sum Y_k \delta_{kj}} \sum y_k = a_i \frac{y_i \sum y_k}{\sum y_k \delta_{kj}} \delta_{ij}$$

据此可知, (1)式由三部分组成: a_i 是模型的待定系数, 对一个特定模型来说, a_i 具有确定的数值, 它规定了每个变量对因变量贡献的相对大小, 而 $Y_i \sum Y_k / \sum Y_k \delta_{kj}$ 是描述同一变量中非“0”数据相对重要性的一个指标, 对 a_i 也有一定的修正作用, 它代表了各单元由于变量组合的差异对相应的 a_i 的校正值。如果把 a_i 作“定权”, 那么第二项则有“变权”的含义, 它只能确定某个单元中该变量所起的作用, 但作用的属性(积极的或消极的)仅取决于前面的系数 a_i ; δ_{ij} 则是原始的定性变量, 它决定在单元内该变量能否作贡献。在整个变量转换过程中, 用 Y_i 来对 δ_{ij} 进行加权乃是区分同一变量中相同 δ_{ij} 值的关键。

将(2)式代入(1)式后, 两边分别对 i 求和得: $\sum Y_i = \sum Y_k \sum a_i \sum Y_i \delta_{ij} / \sum Y_k \delta_{kj} + \sum s_i$

最后解的结果为 $\sum Y_i = 1$ 。

若设 $Y = (y_1, y_2, y_3, \dots, y_n)$; $U_{ij} = Y_i \delta_{ij} \sum Y_k / \sum Y_k \delta_{kj}$, 那么 $A = (a_1, a_2, a_3, \dots, a_m)'$ 的估计值用以下矩阵的形式表达:

$$A = (U'U)U'Y \quad (3)$$

式中: Y —因变量; U —自变量; A —回归方程的系数; U' —矩阵的转置。

经上述变换后, 改善了原始数据矩阵的结构, 同时也使模型中各类变量处于相同的信息等级。即加强了因变量与自变量的联系, 提高模型的复相关系数, 且十分接近于1。值得强调的是: 这仅是一种强化加权的方法, 并不是说因变量和自变量间确实存在着这种相关关系, 模型反映的也可能是一种假相关, 但它并没有改变原有的相关趋势。据此可知, 在前述的两个假设条件下, 这种变换是合理的, 它可以提高模型的拟合程度, 有效地克服建模过程中可能出现的“病态”方程和预测的奇异值。

用上述方法建立的研究区内铜预测储量估算模型和拟合度达99.9%, 估算获得的结果比传统回归预测模型更能为地质人员接受, 获得较好的预测效果。

4 非先验约束模型

非先验约束模型是通过变量间相互关系的表征来突出成矿信息, 新疆阿勒泰侵入岩区金、铜多金属成矿区找矿靶区优选使用的是这类模型。

阿勒泰侵入岩区位于天山—兴安褶皱系西端额尔齐斯大断裂以北。该区地质构造复杂, 加里东期—海西期的岩浆侵入和火山活动频繁, 并形成了相应的成矿系列, 主要是与火山和岩浆热液活动有关的金、铜、铅—锌和与大规模花岗岩化有成因联系的伟晶岩型矿床, 从而使其成为一个重要的有色金属和稀有金属成矿远景区。通过区域成矿规律和成矿系列的研究, 总结了控矿因素和找矿标志, 建立成矿模式和找矿模型, 对全区内的金、铜、铅—锌等矿床的远景作了预测, 圈定了各类矿床的找矿靶区, 进而作了优选。

靶区优选是指在一定的预测准则条件下, 将靶区按成矿有利度或含矿概率进行排序的过程。在现有的数学地质方法中, 有利度通常是指特征分析模型最大特征根所对应的特征向量。它是有一定的地质意义, 而且还有比较直观的表达方式。但有利度的概念也可引伸到因子分析方法中。

因子分析主要特点在于它可用多个意义不同的因子来区分不同矿床模型之间的差异。理论上每个因子都可解释成一个地质作用, 几个因子反映一个地质过程。也就是说, 地质作用在这被抽象分解成因子, 用因子来表示成矿作用的某些特征和成因属性。如在对阿勒泰地区找矿靶区数据的因子分析结果中, 与阿舍勒型铜多金属型矿床(火山热液型)有成因联系的中基性—中酸性火

山岩, 多次活动的北西向断裂, 中上泥盆统至下石炭统层位, 地表铁帽氧化带、铜、铅、锌异常分布和黄铜矿、方铅矿、闪锌矿的矿化特征及矿床的成因参数可用 F_{13} 、 F_1 、 F_4 、 F_{16} 、 F_5 五个因子表征。当然该类矿床的其它一些特征可用另外的因子表达。

地质变量经多维空间旋转后, 用比原来少得多的因子作近似模拟, 这是一个信息浓缩过程。多个因子的存在为区分不同靶区的特征提供了含义丰富的新变量, 但怎样用这些新变量来反映不同矿床模型的成矿有利度呢? 即如何根据因子组合特征提取不同的成矿信息? 这里仍然要涉及强化的概念。解决这些问题的基本理论是在相近地质条件下形成或有成因联系的一组矿床, 尽管其空间位置有所不同, 但它们的控矿条件和找矿标志仍有很大的相似性^[5], 如地表铁帽氧化带都可以作为多拉纳萨依金矿和阿舍勒铜多金属矿床的找矿标志。如果用因子来表达, 有些不同类型的矿床在某些因子上可能仍有相近的比值, 而在另一些因子上, 它们的取值则不同或相反。因此, 对不同类型的矿床, 不仅要研究它们的差异之处, 也要注意彼此间的共同点。只有这样, 才能本质上把握矿床的重要特征和有机联系。

既然因子能反映矿床的基本特征和成因属性, 我们就可将找矿靶区与矿床类型间的类比转化成各有关因子之间的对比。由于因子已不同于原始变量, 它们具有相同的量纲分布类型, 能够对其进行各种线性组合。为便于定量对比, 我们按数学原理和可能反映成矿有利度定量特征的角度出发, 设计了如下找矿靶区有利度的线性模型:

$$F = a_1 f_1 + a_2 f_2 + a_3 f_3 + \dots + a_{20} f_{20} \quad (4)$$

其中: f_k —找矿靶区上在第 k 个因子上的得分, $k=1, 2, 3, \dots, 20$ (在阿勒泰靶区优选原始数据的因子分析结果共取20个因子);

a_k —第 k 个因子的强化系数, 其值暂定为“ -1 ”, “ 0 ”, “ 1 ”, 以此表示第 k 个因子在所选定的模型中是否起作用, 起什么样的作用。

用20个地质因子经过新的线性组合后, 在原来的样品空间中形成了一个新的变量。从模型的地质意义出发, 新变量是各个有利信息的叠合。它能否反映成矿有利信息, 就取决于强化系数的选择和确定。

在阐明各因子地质意义的基础上, 重点分析典型矿床所在靶区的因子得分形式及其强度。为对该方法进行探索, 选定区内不同矿种的4种类型共10个矿床所在的靶区作为模型区, 它们是4个金矿床, 3个铜矿床, 1个铜-镍矿床, 2个铅-锌矿床。在地质意义明确的因子上, 得分(绝对值)较高的靶区一般可以认为有利因素存在。诸多的有利因素组合起来, 便构成了靶区成矿的有利性。如果选择对模型区影响较大的因子作为有利标志, 那么在预测区和模型区的类比过程中, 可以突出关键的控矿因素, 克服或抑制次要因素的干扰。为此在确定强化系数时, 对在模型区取值很小的因子赋“ 0 ”, 而对于较大正值者赋“ 1 ”, 较大负值者赋“ -1 ”。由于这个不同类型矿床所在单元作为已知模型的因子得分和趋向, 因此每个模型的强化系数构成也有较大的差别。如多拉纳萨依模型(破碎蚀变岩型)和恰本型(蚀变花岗岩型)金矿的强化系数构成分别为:

$$(0, 0, 0, 1, 0, -1, 0, 0, 0, 0, 1, 1, 0, 0, 1, 0, 1, 0, 1, 0)$$

$$(-1, 0, 0, -1, 0, 1, 1, 1, 1, 0, 0, 0, 1, -1, -1, 0, -1, 1, 0, -1)$$

从以上二个已知金矿模型的强化系数的构置情况的实际是: 虽然都属金矿床, 但由于类型和成因机理的差异, 两者的因子得分强度和趋向均有明显的差异。用强化系数可将两类不同性质的找矿靶区及其有利度得到区分。

用同样原则对阿勒泰地区10个已知矿床模型的因子信息强化系数作出合理构置, 然后分别用10个模型各自的强化系数按(4)式计算出全部靶区的总的有利度 F 。这样每个找矿靶区可拥有10个有利度。最后将这10个有利度从大到小进行排序, 并按有利度的大小分级, 再将大部分有利

度太小的找矿靶区分别予以删除。计算结果表明，大部分矿床因子模型之间差别是显著的。特别是原生金矿和多金属矿床各模型的A级和B级找矿靶区基本上没有相互穿插，而C级找矿靶区的互容现象比较明显。反映出这些模型的成矿条件之间既有差别，又有联系。这一方法优选结果反映的实质是区内各已知的成型矿床均已进入各自的A级找矿靶区，其有利度值较高，从理论上和实际计算结果证实是有效。在地质-找矿的实践中也已证实这一方法是有效的。区内的阿舍勒铜-锌矿床是成型的大型矿床，用它构置的强化系数计算其它找矿的有利度数值表明区内有四个单元可能赋存有与阿舍勒型矿床相似的远景单元，后来经二年的野外地质工作证实，已在仅次于阿舍勒模型有利度的68号单元的地表和深部（钻探）见到较好的铜-锌矿床，钻探证实已成型，并还在不断扩大。

以上是两个对数学或预测模型进行强化的例子。强化是对找矿靶区进行优化的一种途径，它可以用多种方法来实施，应用强化系数仅是方法之一。对于不同类型矿床、或不同地区可采用不同的方法。但强化的目的是改善预测模型的贴切程度，提高矿床预测成果的可靠性，有利于地质-找矿工作的部署和提高地质-找矿效益。

参 考 文 献

- 1 朱裕生. 矿产资源评价方法学导论. 北京: 地质出版社, 1984. 71页.
- 2 张炳熹, 陈毓川. 成矿模式与成矿系列. 当代地质科学动向. 北京: 地质出版社, 1987. 148页.
- 3 浦志伟. 资源信息处理过程中定性数据的QMBD模型及其应用. 地学探索, 1988, (1): 60.
- 4 朱裕生. 地质数据的预处理方法. 中国地质科学院矿床地质研究所刊, 1986, (1): 50.
- 5 朱裕生. 指导成矿预测的地质理论——成矿学理论和应用. 中国地质, 1988, (4): 7.

THE ABSTRACTION AND ENHANCEMENT OF MINERALIZATION INFORMATION IN PROGNOSIS OF MINERAL RESOURCES

Pu Zhiwei and Zhu Yusheng

(Chinese Academy of Geological Sciences, Beijing)

Abstract

The extraction, concentration and enhancement of information about mineralization are a key technological problem in prognosis of mineral resources. It calls for consciously meddling in the construction of the metallogenic model, giving prominence to the information related to mineralization and prognosis, controlling or eliminating some information that has no obvious mineralization significance or interferes with prognosis, promoting the prognostic model to approximate to what can reflect the mineralization information and promoting the coincidence between the model and the actual geological setting of mineral deposits according to our knowledge of the metallogeny of mineral deposits and adhering to certain criteria for prognosis. Proceeding from the study of the metallogenic model and ore-controlling geological indications, this paper suggests that the new technique of information enhancement should be applied in prognosis of mineral potential. Some examples of the enhancement method for prioriconfined models and non-

priori-confined models are given in the paper. Some mineral deposits of commercial value have been found in the ore target areas determined by using this method; good ore prospecting results have been achieved. The authors expound the method, steps and geological prerequisites of application of the "information enhancement" technique. Practice has proved that the enhancement of mineralization information is not only conducive to extracting useful information but also strengthening an organic combination of the metallogenic model with the prognostic model of mineral deposits, thus raising the effectiveness of mineral prospecting.

Key words: mineral information, prognosis of mineral resources, information extraction and enhancement

作 者 简 介

浦志伟, 生于1959年6月9日, 籍贯江苏无锡, 1982年毕业于南京大学地质系, 1986年于中国地科质科学研究生部毕业, 授于硕士。后任中国地质科学院区划室工程师, 中国地质学会青年工作委员会委员。从事成矿预测方法研究。通讯处: 北京百万庄中国地质科学院成矿远景区划室, 邮政编码: 100037。