

改进的加权证据权模型及其在个旧锡铜矿产资源预测中的应用

张生元^{1,2}, 成秋明², 张素萍¹, 徐德义^{2,3*}

1. 石家庄经济学院资源与环境工程研究所, 河北石家庄 050031

2. 中国地质大学地质过程与矿产资源国家重点实验室, 湖北武汉 430074

3. 中国地质大学经济管理学院, 湖北武汉 430074

摘要: 为了探讨新的加权系数估计方法对于消除或减弱证据层不满足条件独立性假设时对预测结果的影响, 对加权证据权模型的加权系数估计方法进行了新的探讨, 尝试用顺序估计法估计加权系数. 加权系数的顺序估计法是将加权证据权模型与基于模糊预测对象的证据权模型相结合, 将证据层按照一定顺序逐步加入到加权证据权模型中, 在加入到模型的过程中依次用已经获得的后验概率作为模糊训练层对证据层加入到模型的顺序进行修正, 并通过条件相关系数的方法估计加权系数. 分别以1组多元正态分布模拟数据和个旧锡铜多金属矿产资源预测为例, 比较了多种模型的后验概率, 结果表明加权证据权模型对减弱证据层不满足条件独立性假设所产生的影响是有效的.

关键词: 加权证据权; 模糊证据权; 顺序估计; 资源预测; 数学模型; 数学地质; 成矿预测.

中图分类号: P628

文章编号: 1000-2383(2012)06-1175-08

收稿日期: 2012-07-19

Improvement of Weighted Weights of Evidence and Its Applications in Sn-Cu Mineral Potential Mapping in Gejiu, Yunnan Province, China

ZHANG Sheng-yuan^{1,2}, CHENG Qiu-ming², ZHANG Su-ping¹, XU De-yi^{2,3*}

1. Institute of Natural Resources and Environmental Engineering, Shijiazhuang University of Economics, Shijiazhuang 050031, China

2. State Key Laboratory of Geology Processes and Mineral Resources, China University of Geosciences, Wuhan 430074, China

3. School of Economics and Management, China University of Geosciences, Wuhan 430074, China

Abstract: This paper proposes a new method to estimate weighting coefficients in weighted weights of evidence (WWofE) in order to reduce the influence of correlation among evidence layers when hypothesis of conditional independence is not true. Sequential estimation method in WWofE combines WWofE and dual weights of evidence (DWofE) to gradually add evidence layers with specific sequence to WWofE model. In the process, the posterior probability obtained by former evidence layers is considered as fuzzy training layer in DWofE to correct the sequence of layers. The weighting coefficients are estimated by using conditional correlation coefficient. As a case study, a group data generated by four variable normal distribution and Sn-Cu mineral resources assessment in Gejiu, Yunnan, southwestern China are used. The results show that WWofE model is effective to reduce the influence of correlations among evidence layers.

Key words: weighted weights of evidence; weighing weights of evidence; sequential estimation method; mineral resources assessment; mathematical model; mathematical geology; ore prospecting.

证据权方法是一种贝叶斯方法, 该方法最初被用于医疗诊断中, 20世纪80年代, 加拿大 Agter-

berg and Bonham-Carter 等数学地质学家对此方法进行了修改和发展, 将其应用于矿产预测 (Agter-

基金项目: 国家自然科学基金 (Nos. 41172299, 40972205); 地质调查项目 (No. 1212011085468); 地质过程与矿产资源国家重点实验室开放课题 (No. GPMR200803).

作者简介: 张生元 (1961-), 男, 博士, 教授, 主要从事数学地质以及矿产资源定量评价方法和开发科研和教学工作.

* **通讯作者:** 徐德义, E-mail: xdy@cug.edu.cn

berg *et al.*, 1989; Agterberg, 1989). 证据权模型经过 20 多年的发展, 无论是在应用, 还是在模型改进以及软件开发方面都取得了很大进展. 在应用方面, 证据权模型已被国内外广泛应用到矿产资源定量评价、滑坡敏感性评价、地质灾害和生态学等多个领域.

尽管证据权模型在国内外多个领域得到广泛应用, 但证据权模型要求证据层必须满足条件独立性, 而证据层(地质变量)是由成矿预测模型决定的, 这些地质变量之间客观上存在一定的相关关系, 因此, 要让这些证据层满足条件独立性是非常困难的, 甚至是不可能的. 不能满足条件独立性的结果将造成后验概率偏大或者偏小, 而且由于后验概率估计不准确, 会造成其估计未知矿床数时偏差较大, 这制约了证据权模型在矿产资源评价中对资源量的有效估计.

证据权模型的条件独立性假设是证据权模型应用中的难题, 长期以来为了最大程度上满足条件独立性, 人们不得不通过调整地质变量的分类或减少模型中的地质变量来实现, 然而某些被去掉的变量可能与成矿有密切关系, 这导致有用信息浪费. 这方面的最新研究进展包括: 有学者提出的模糊证据权模型和 A-C 模型 (Cheng and Agterberg, 1999; Agterberg and Cheng, 2002). Journal(2002)提出了一种新的地质统计学模型 (Krishnan *et al.*, 2005; Krishnan, 2008), 该模型为改进证据权模型在条件独立性方面的限制提供了新思路, Polyakova and Journal(2007)在 tau 模型的基础上提出了 nu 模型. Cheng(2008, 2012)提出了综合地质、地球化学、地球物理等控矿要素预测矿产资源后验概率的多重级联模型(奇异性证据权模型), 放宽了证据权法使用的前提条件, 在一定程度上提高了证据权模型预测效果; 张生元等(2006)提出了基于模糊预测对象的证据权模型, 解决了证据权模型模糊对象的预测问题. 张生元等(2009)提出了加权证据权模型, 通过加权系数降低证据层之间存在的空间相关性对后验概率的影响, 为解决证据层不独立的难题提供了新思路和新方法. 但加权系数如何赋值才能提高证据权模型的预测精度, 引入加权系数后是否会引入新的误差尚需进一步深入的研究. Deng(2009)提出了条件相依的证据权模型(CDAWE), 该模型通过与逻辑回归模型相结合对证据权重进行修正, 极大地减弱了证据图层不满足条件独立性假设对预测的影响, 但 Schaabben and Boogaart(2011)的实例研究表

明, 他的模型存在致命缺陷, 不能推广到多个变量的情况. Agterberg(2011)将证据权模型和逻辑回归方法结合, 提出了一种修正证据权模型.

1 tau 模型

Journal(2002)提出了一种 tau 模型, Krishnan *et al.* (2005)和 Krishnan(2008)对该模型进行了完善, 模型如下:

假设 A_1, A_2, \dots, A_n 是 n 个具有多分类数据的证据图层, D 是预测事件. 定义距离为:

$$x = \frac{P(\bar{D} | A_1, \dots, A_n)}{P(D | A_1, \dots, A_n)}, \quad (1)$$

$$x_i = \frac{P(\bar{D} | A_i)}{P(D | A_i)}, i = 1, \dots, n, \quad (2)$$

$$x_0 = \frac{P(\bar{D})}{P(D)}, \quad (3)$$

公式(1)~(3)的取值范围为 $[0, \infty)$, 其定义为几率(odds)的倒数.

定义 tau 模型为:

$$\frac{x}{x_0} = \prod_{i=1}^n (x_i)^{\tau_i}. \quad (4)$$

则,

$$P(D | A_1, \dots, A_n) = \frac{1}{1+x} \in [0, 1],$$

式中: $\tau_i \in (-\infty, \infty)$, 当 $P(D | A_i) = 1$ 或 0 时, $P(D | A_1, \dots, A_n) = 1$ 或 0 . 对公式(4)取对数得到:

$$\log(x) - \log(x_0) = \sum \tau_i [\log(x_i) - \log(x_0)]. \quad (5)$$

2 加权证据权模型

与 tau 模型相同, 假设 A_1, A_2, \dots, A_n 是 n 个具有多分类数据的证据图层, D 是预测事件. 普通证据权模型为:

$$\text{logit}(D | A_1 A_2 \dots A_n) = W_0 + W_1 + \dots + W_n, \quad (6)$$

式中: W_1, W_2, \dots, W_n 为证据图层的证据权重. 其中:

$$W_0 = \log \frac{P(D)}{P(\bar{D})}, \quad (7)$$

$$W_i = \log \frac{P(A_i | D)}{P(A_i | \bar{D})}, i = 1, \dots, n. \quad (8)$$

上述关系的成立要求证据图层满足条件独立性假设. 如果证据图层不满足条件独立性条件, 则公式

(6)不成立. 注意到公式(6)的左边与 tau 模型有如下关系:

$$\begin{aligned} &\text{logit}(P(D | A_1, \dots, A_n)) = \\ &\log \frac{P(D | A_1, \dots, A_n)}{1 - P(D | A_1, \dots, A_n)} = \frac{1}{x} \end{aligned}$$

而预测事件 D 与逆事件 \bar{D} 的对称性,定义加权证据权模型(张生元等,2009),其 $\log it$ 模型为:

$$\text{logit}(D | A_1 A_2 \dots A_n) = W_0 + \tau_1 W_1 + \dots + \tau_n W_n \tag{9}$$

式中: W_1, W_2, \dots, W_n 为证据图层的证据权重. $\tau_1, \tau_2, \dots, \tau_n$ 为加权证据权模型的加权系数, $\tau_i \in (-\infty, +\infty)$. 从公式(9)可以看出,如果加权系数 τ_i ($i=1, \dots, n$)都等于 1,则加权证据权模型就是普通的证据权模型.

加权证据权模型用乘积形式表示为:

$$\begin{aligned} &\frac{P(D | A_1 A_2 \dots A_n)}{P(\bar{D} | A_1 A_2 \dots A_n)} = \\ &\frac{P(D)}{P(\bar{D})} \left(\frac{P(A_1 | D)}{P(A_1 | \bar{D})} \right)^{\tau_1} \left(\frac{P(A_2 | D)}{P(A_2 | \bar{D})} \right)^{\tau_2} \dots \left(\frac{P(A_n | D)}{P(A_n | \bar{D})} \right)^{\tau_n} \end{aligned} \tag{10}$$

在公式(9)中证据图层的证据权重与公式(6)的证据权重相同,都由公式(7)和公式(8)定义. 因此,在加权证据权模型中,如何确定加权系数 $\tau_1, \tau_2, \dots, \tau_n$ 是加权证据权模型的关键.

3 加权系数的计算

3.1 顺序加权系数估计法

设 $A_1^{(s)} A_2^{(s)} \dots A_n^{(s)}$ 是 n 个证据图层 A_1, A_2, \dots, A_n 的任意一种排列顺序,则对于乘积模型公式(10)有:

$$\begin{aligned} &\frac{P(D | A_1 A_2 \dots A_n)}{P(\bar{D} | A_1 A_2 \dots A_n)} = \frac{P(D | A_1^{(s)} A_2^{(s)} \dots A_n^{(s)})}{P(\bar{D} | A_1^{(s)} A_2^{(s)} \dots A_n^{(s)})} = \\ &\frac{P(D)}{P(\bar{D})} \left(\frac{P(A_1 | D)}{P(A_1 | \bar{D})} \right)^{\tau_1} \left(\frac{P(A_2 | D)}{P(A_2 | \bar{D})} \right)^{\tau_2} \dots \left(\frac{P(A_n | D)}{P(A_n | \bar{D})} \right)^{\tau_n} \end{aligned}$$

注意到:

$$\begin{aligned} &\frac{P(D | A_1^{(s)} A_2^{(s)} \dots A_n^{(s)})}{P(\bar{D} | A_1^{(s)} A_2^{(s)} \dots A_n^{(s)})} = \\ &\frac{P(D)}{P(\bar{D})} \left(\frac{P(A_1^{(s)} | D)}{P(A_1^{(s)} | \bar{D})} \right) \left(\frac{P(A_2^{(s)} | DA_1^{(s)})}{P(A_2^{(s)} | \bar{D}A_1^{(s)})} \right) \dots \left(\frac{P(A_n^{(s)} | DA_1^{(s)} \dots A_{i-1}^{(s)})}{P(A_n^{(s)} | \bar{D}A_1^{(s)} \dots A_{i-1}^{(s)})} \right) \end{aligned}$$

令:

$$\tau_i^{(s)} = \frac{\log \frac{P(A_1^{(s)} A_2^{(s)} \dots A_{i-1}^{(s)} | D)}{P(A_1^{(s)} A_2^{(s)} \dots A_{i-1}^{(s)} | \bar{D})}}{\log \frac{P(A_i^{(s)} | D)}{P(A_i^{(s)} | \bar{D})}}, i = 2, \dots, n, \tag{11}$$

则上式可写为:

$$\begin{aligned} &\frac{P(D | A_1^{(s)} A_2^{(s)} \dots A_n^{(s)})}{P(\bar{D} | A_1^{(s)} A_2^{(s)} \dots A_n^{(s)})} = \\ &\frac{P(D)}{P(\bar{D})} \left(\frac{P(A_1^{(s)} | D)}{P(A_1^{(s)} | \bar{D})} \right)^{\tau_1^{(s)}} \left(\frac{P(A_2^{(s)} | D)}{P(A_2^{(s)} | \bar{D})} \right)^{\tau_2^{(s)}} \dots \left(\frac{P(A_n^{(s)} | D)}{P(A_n^{(s)} | \bar{D})} \right)^{\tau_n^{(s)}} \end{aligned}$$

因此,公式(11)可以看作是加权系数的估计,但也可以看出, n 个证据图层不同的排列顺序, $\tau_i^{(s)}$ 的值是不一样的. 也就是说用公式(11)作为加权系数的估计,受证据层的排列顺序影响,同时也要注意 $\tau_i^{(s)}$ 的估计与 $A_1^{(s)} A_2^{(s)} \dots A_n^{(s)}$ 的取值有关.

加权系数 $\tau_i^{(s)}$ 的性质有以下 5 点:

$$(1) \text{ 当 } \tau_i^{(s)} = 1, \text{ 则 } \frac{P(A_i^{(s)} | DA_1^{(s)} \dots A_{i-1}^{(s)})}{P(A_i^{(s)} | \bar{D}A_1^{(s)} \dots A_{i-1}^{(s)})} =$$

$\frac{P(A_i^{(s)} | D)}{P(A_i^{(s)} | \bar{D})}$,也就是说 $(A_1^{(s)}, \dots, A_{i-1}^{(s)})$ 对于 W_i 的计算没有影响,特别是当证据图层满足条件独立时, $\tau_i^{(s)} = 1$,因此,普通证据权项目是加权证据权模型的特例,但是 $\tau_i^{(s)} = 1$ 并不能说证据图层满足条件独立性.

(2)当 $\tau_i^{(s)} = 0$ 时,则 $P(A_i^{(s)} | DA_1^{(s)} \dots A_{i-1}^{(s)}) = P(A_i^{(s)} | \bar{D}A_1^{(s)} \dots A_{i-1}^{(s)})$,说明证据图层 A_i 对预测事件 D 及其逆 \bar{D} 的出现没有作用,也就是说该证据图层就 D 的预测而言是无效的.

(3)当 $0 < \tau_i^{(s)} < 1, \tau_i^{(s)} > 1$ 时, $\log\left(\frac{P(A_i^{(s)} | DA_1^{(s)} \dots A_{i-1}^{(s)})}{P(A_i^{(s)} | \bar{D}A_1^{(s)} \dots A_{i-1}^{(s)})}\right) > \log\left(\frac{P(A_i^{(s)} | D)}{P(A_i^{(s)} | \bar{D})}\right) = W_i^{(s)}$,说明在 $(A_1^{(s)}, \dots, A_{i-1}^{(s)})$ 出现的条件下, A_i 对事件 D 的预测敏感度提高.

(d)当 $\tau_i^{(s)} < 0$ 时, $\log\left(\frac{P(A_i^{(s)} | DA_1^{(s)} \dots A_{i-1}^{(s)})}{P(A_i^{(s)} | \bar{D}A_1^{(s)} \dots A_{i-1}^{(s)})}\right) < \log\left(\frac{P(A_i^{(s)} | D)}{P(A_i^{(s)} | \bar{D})}\right) = W_i^{(s)}$,说明在 $(A_1^{(s)} \dots A_{i-1}^{(s)})$ 出现的条件下, A_i 对事件 D 的预测敏感度降低.

(5)当 $\tau_i^{(s)} = \infty$ 或 $-\infty$ 时,则说明 $P(A_i^{(s)} | DA_1^{(s)} \dots A_{i-1}^{(s)}) = 0$ 或 $P(A_i^{(s)} | \bar{D}A_1^{(s)} \dots A_{i-1}^{(s)}) = 1$.

3.2 顺序加权系数相关系数估计法

由上可知, A_1, A_2, \dots, A_n 的不同排列将导致不同的预测结果,逐步校正方法的思想就是首先寻找 A_1, A_2, \dots, A_n 的最佳排列,然后逐步对加权证据权模型进行校正(Krishnan,2008). 基本步骤如下:

按照各证据图层的熵从大到小的顺序排列 A_1, A_2, \dots, A_n . 不妨设为 A_1, A_2, \dots, A_n . 按照如下方法估计加权系数,

$$\tau_1 = 1, \tau_2 = 1 - \rho_{(A_1, A_2 | D)}^2, \tag{12}$$

式中: $\rho_{(A_1, A_2 | D)}$ 表示证据层 A_1 与 A_2 关于 D 的条件

相关系数,如果 A_1 与 A_2 关于 D 是条件独立的,则 $\rho_{(A_1, A_2|D)}$, 否则, $\rho_{(A_1, A_2|D)}$ 不等于 0, τ_2 就是试图降低 2 个证据层条件相关产生的影响。

$$\tau_3 = 1 - \rho_{(A_3, A_1 A_2|D)}^2, \quad (13)$$

式中: $\rho_{(A_3, A_1 A_2|D)}$ 表示证据层 A_3 与 (A_1, A_2) 关于 D 的条件相关系数. 如果 A_3 与 (A_1, A_2) 关于 D 是条件独立的,则 $\rho_{(A_3, A_1 A_2|D)} = 0$, 否则 $\rho_{(A_3, A_1 A_2|D)}$ 不等于 0, τ_3 就是试图降低 (A_1, A_2) 与 A_3 条件相关产生的影响。

$$\tau_i = 1 - \rho_{(A_i, A_1 A_2 \dots A_{i-1}|D)}^2, i = 4, \dots, n, \quad (14)$$

式中: $\rho_{(A_i, A_1 A_2 \dots A_{i-1}|D)}$ 表示证据层 A_i 与前 $i-1$ 个证据层 (A_1, \dots, A_{i-1}) 关于 D 的条件相关系数. 由此可以估计出所有加权系数的值。

为了确定证据图层的最佳顺序,一种方法可以用普通证据权模型的显著性指标 t -(student) 统计量进行确定,首先计算各个证据图层的 t_i , 然后按照从大到小的顺序进行排列,其顺序作为证据图层的顺序. 另一种方法可以用条件信息熵的方法:

$$H(A_i) = - \sum \log(P(D | A_i)P(D | \bar{A}_i)),$$

$i = 1, \dots, n.$

将 $H(A_i)$ 按照从大到小的顺序进行排列,其顺序作为证据图层的顺序。

3.3 顺序加权系数相关系数估计法的基本步骤

上述方法确定证据图层的顺序只考虑了单个图层与 D 的关系,并没有考虑图层之间关于预测事件存在交互效应. 事实上当加入某个图层以后,剩余图层对预测事件 D 的作用会受到前面已经加入到证据图层的影响,因此,在将剩余图层顺序加入到模型时不仅要考虑剩余图层对预测事件的影响,还要考虑前面已经加入到模型中的图层对要加入图层的影响. 逐步顺序加权的基本步骤如下:

(1) 确定第 1 个被加入的图层. 首先计算各个证据图层的 t - 统计量 $t_i (i=1, \dots, n)$, 然后取 t_i 最大值所对应的图层作为下一个加入模型的图层,不失一般性记为 A_1 , 则用证据权模型计算 $P(D|A_1)$;

(2) 确定剩余图层的加入顺序. 将公式(1)计算的后验概率 $P(D|A_1)$ 作为模糊训练层,用双模糊证据权模型计算剩余图层的 t - 统计量 t_i , 然后取 t_i 最大值对应的图层作为下一个加入模型的图层,不妨记为 A_2 ;

(3) 计算加权系数 τ_2 . 由公式(12), 首先计算 $\rho(A_1 A_2 | D)$, 再计算 τ_2 ;

(4) 用加权证据权模型公式(9)计算后验概率

$P(D|A_1 A_2)$;

(5) 按照公式(2)到公式(4)相同的步骤,选择下一个加入模型的图层,计算加权系数以及后验概率,直到所有图层加入到模型中。

3.4 无顺序相关系数估计法

从加权证据权模型的公式(9)可以看出,加权系数与证据图层的顺序是无关系的,为了获得加权系数与证据图层顺序无关的估计,仍然用相关系数方法. 定义:

$$\tau_i = 1 - \rho(A_i, \prod_{j \neq i, 1 \leq j \leq n} A_j | D)^2, i = 1, \dots, n,$$

式中: $\rho(A_i, \prod_{j \neq i, 1 \leq j \leq n} A_j | D)$ 为在 D 的条件下 A_i 与 $\prod_{j \neq i, 1 \leq j \leq n} A_j$ 的条件相关系数. 在此定义下,如果模型满足条件独立性假设,则 $\tau_i = 1$, 此时的加权证据权模型就是证据权模型,如果某证据层与模型中的其他证据层完全相关,即 $|\rho(A_i, \prod_{j \neq i, 1 \leq j \leq n} A_j | D)| = 1$, 则 $\tau_i = 0$, 也就是说该证据层在模型中没有作用。

3.5 条件相关系数的计算

3.5.1 后验概率近似计算法 一般来说,条件相关系数 $\rho(A_i, \prod_{j \neq i, 1 \leq j \leq n} A_j | D)$ 或 $\rho_{(A_i, A_1, \dots, A_{i-1}|D)}$ ($i = 2, \dots, n$) 的严格计算不是很容易,这里涉及到 1 个变量和 1 组变量的条件相关系数,张生元等(2009)建议使用如下方法计算条件相关系数:

① 用模糊证据权模型计算 A_i 的每个子区域的显著性,并按照其显著性从大到小进行排序,以排序值代替其分类值生成新的图层;

② 首先将其他证据层进行叠加,将叠加后的图层在模糊证据权模型下计算每个子区域的显著性,并按照从大到小的顺序进行排列,以排序值代替其分类值生成新的图层;

③ 把模糊证据权模型得到的后验概率除以其和,使其满足概率的定义,将其看作是①和②生成的 2 个图层在 D 条件下的联合概率分布的近似,在此基础上计算这 2 个图层的相关系数,用此相关系数作为 $\rho(A_i, \prod_{j \neq i, 1 \leq j \leq n} A_j | D)$ 的近似值。

3.5.2 条件相关系数 当 $i > 2$ 时,需要计算 $\rho_{(A_i, A_1, \dots, A_{i-1}|D)}$, 涉及 1 个证据图层 A_i 与多个证据图层 (A_1, \dots, A_{i-1}) 的条件相关系数,通常这个条件相关系数计算比较困难. 为了简化计算,首先将 (A_1, \dots, A_{i-1}) 进行叠加生成 1 个新的图层 B , 利用如下公式计算条件相关系数:

$$\rho_{(A_i, A_1, \dots, A_{i-1}|D)} \approx \rho_{(A_i, B)}$$

$$\rho_{(A_i, B|D)} = \frac{\rho_{(A_i, B)} - \rho_{(A_i, D)} \rho_{(B, D)}}{\sqrt{1 - \rho_{(A_i, B)}^2} \sqrt{1 - \rho_{(B, D)}^2}}.$$

4 模拟数据试验和应用实例

4.1 正态分布模拟数据试验

首先用正态分布的模拟数据进行试验,其目的是检验当证据变量不满足条件独立性假设时,对后验概率估计的影响,以及加权证据权模型是否对普通证据权模型有所改善.利用 Matlab 的 mvnrnd 函数从 4 元正态分布 $N(\mu, \Sigma)$ 随机抽取 10 000 个样本,其中, $\mu=[2, 5, 8, 11]$ 为总体均值向量, Σ 是总体方差矩阵:

$$\Sigma = \begin{bmatrix} 1 & 0.8 & 0.7 & 0.9 \\ 0.8 & 1 & 0.7 & 0.85 \\ 0.7 & 0.7 & 1 & 8 \\ 0.9 & 0.85 & 0.8 & 1 \end{bmatrix},$$

对应的随机变量表示为 (X_1, X_2, X_3, Y) , 其中 X_1, X_2, X_3 作为证据变量, Y 预测变量, 从总体方差矩阵 Σ 可以看出, 其就是相关系数矩阵, 而且各个变量之间存在相关性. 为了应用证据权模型, 将上述 4 个正态随机变量分别进行二值化处理,

$$A = \begin{cases} 1, X_1 \geq 2 \\ 0, \text{其他} \end{cases}, B = \begin{cases} 1, X_2 \geq 5 \\ 0, \text{其他} \end{cases},$$

$$C = \begin{cases} 1, X_3 \geq 8 \\ 0, \text{其他} \end{cases}, D = \begin{cases} 1, Y \geq 11 \\ 0, \text{其他} \end{cases}.$$

设 $P(A=1)=P(B=1)=P(C=1)=P(D=1)=0.5$, 对 10 000 个已抽取的样本按照相同的方法进行二值化, 并分别计算样本相关系数 Σ_s 和二值化样本相关系数 Σ_b 如下:

$$\Sigma_s = \begin{bmatrix} 1 & 0.8018 & 0.6985 & 0.9033 \\ 0.8018 & 1 & 0.6975 & 0.8504 \\ 0.6985 & 0.6975 & 1 & 0.7956 \\ 0.9033 & 0.8504 & 0.7956 & 1 \end{bmatrix},$$

$$\Sigma_b = \begin{bmatrix} 1 & 0.6022 & 0.4904 & 0.7210 \\ 0.6022 & 1 & 0.4842 & 0.6492 \\ 0.4904 & 0.4842 & 1 & 0.5751 \\ 0.7210 & 0.6492 & 0.5751 & 1 \end{bmatrix}.$$

从这 2 个样本相关系数可以看出, 连续变量的二值化会使得变量之间的相关关系发生改变. 为了进一步比较不同证据权模型对后验概率估计的影响, 用 Matlab 的 mvncdf 函数分别计算了 A, B, C 不同组合对应的后验概率 $P(D=1|ABC)$ 的理论值, 并对这些后验概率理论值按照从大到小的顺序进行排列. 用二值化后的样本估计证据权权重, 用普通证据权模型计算相应的后验概率, 对其从大到小排序(表 1).

为了确定证据变量和预测变量之间关系, 利用二值化样本计算预测变量 D 与证据变量 A, B, C 的偏相关系数分别是 0.489 8, 0.322 9, 0.294 6, 并按照偏相关系数从大到小顺序确定 A, B, C 加入到 WWofE 的顺序为 $A \rightarrow B \rightarrow C$. 由公式(12)和公式(13)计算得, $\tau_1=1, \tau_2=1-\rho_{(B,A|D)}^2=0.9352, \tau_3=1-\rho_{(C,AB|D)}^2=0.9600$, 用加权证据权模型计算相应的后验概率, 对其从大到小排序(表 1). 表 1 最后一列是不同模型后验概率估计值与后验概率理论值的误差平方和.

从表 1 可以看出:(1)证据权变量关于预测变量不满足条件独立性使得后验概率估计出现偏大或者偏小的问题;(2)尽管 2 种证据权方法估计的后验概率与理论后验概率之间存在误差, 但按其从大到小排列的顺序与理论后验概率的顺序基本一致;(3)将 2 种证据权模型后验概率的估计值与后验概率的理论值进行比较发现, 除 $\bar{A}B\bar{C}$ 和 $A\bar{B}C$ 外, 其他 6 种情况加权证据权模型的后验概率比普通证据权模型的后验概率更接近后验概率的理论值, 加权证据权模型的误差平方和比普通证据权模型的误差平方和小, 说明加权证据权模型与普通证据权模型相比有一定的改善.

4.2 加权证据权模型对个旧锡铜多金属矿产资源的预测

本实例选择的研究区为云南个旧锡铜多金属矿区, 总面积约 1700 km². 云南个旧锡铜多金属矿区

表 1 各种方法计算的后验概率及排序

Table 1 Four posterior probability and their ranks in unique condition

类型	$\bar{A}B\bar{C}$	$A\bar{B}C$	$\bar{A}B\bar{C}$	$AB\bar{C}$	$\bar{A}BC$	$A\bar{B}C$	$\bar{A}BC$	ABC	误差
后验概率理论值	0.034	0.409	0.249	0.765	0.235	0.752	0.590	0.966	
后验概率理论值排序	8	5	6	2	7	3	4	1	
普通证据权后验概率	0.010	0.269	0.176	0.891	0.117	0.835	0.746	0.991	55.100
普通证据权后验概率排序	8	5	6	2	7	3	4	1	
加权证据权后验概率	0.013	0.300	0.169	0.885	0.122	0.841	0.716	0.989	45.600
加权证据权后验概率排序	8	5	6	2	7	3	4	1	

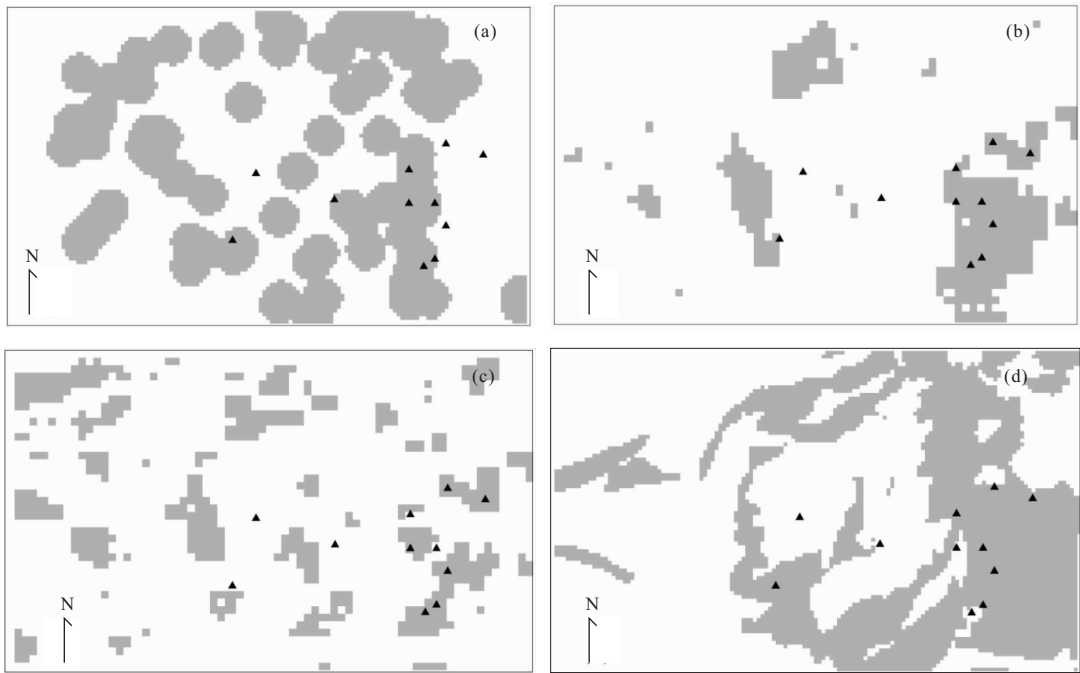


图 1 4 个证据图层二态图层(张生元等,2009)

Fig. 1 Binary maps of four evidence maps

a. 构造交汇点距离 6 km, 白色点表示构造交汇点; b. 采用 S-A 方法分解得到的地球化学综合异常图; c. 采用局部奇异性方法得到的局部地球化学异常图; d. 个旧组地层. 黑色三角形表示 Sn 矿床

位于云南省东南部,是中国滇东南锡矿带上最重要的、规模最大的锡多金属矿集区之一. 为了应用证据权模型,选取研究区内的 11 个矿床构成训练层,4 个图层作为证据层(图 1). 有关证据层的选择和确定参见成秋明等(2009a, 2009b).

4.2.1 逐步加权证据权模型应用 4 个证据层对矿床出现的显著性按照从大到小排列依次是: S-A 方法得到的地球化学综合异常图(图 1b)、局部奇异性方法得到的局部地球化学异常图(图 1c)、个旧组地层(图 1d)和构造交汇点 6 km 缓冲区(图 1a)(张生元等,2009),因此,笔者按照顺序: b→c→d→a, 把 4 个证据层逐步加入到证据权模型中. 加权系数按照逐步顺序估计法,相关系数按照后验概率相关系数法,经计算得 $\tau_1 = 1.000$, $\tau_2 = 0.757$, $\tau_3 = 0.566$, $\tau_4 = 0.999$, 并计算后验概率.

4.2.2 结果比较 为了进行比较,将张生元等(2009)使用相同数据的 2 种模型,模糊证据权方法(I)和无序加权证据权模型(II)所获得的结果直接引入到这里,同时把公式(1)中使用的逐步加权证据权模型记为模型(III). 对上述 3 种方法得到的后验概率在 4 个证据层叠加后的不同子区域上进行分区统计,并按照后验概率从大到小排序(表 2). 用表 2 中后验概率的排序结果计算 3 个后验概率排序的相

表 2 基于各个子区域 3 种证据权模型后验概率从大到小排序

Table 2 The rank of posterior probabilities obtained by using three models in unique condition

子区域	面积单元数	所含矿床数	I	II	III
ABCD	31.6	1	1	1	1
ABC \bar{D}	19.4	2	3	3	2
AB $\bar{C}D$	22.0	3	5	4	5
AB $\bar{C}\bar{D}$	10.3	0	8	8	6
A $\bar{B}CD$	23.6	0	7	6	8
A $\bar{B}\bar{C}\bar{D}$	20.7	0	11	10	10
A $\bar{B}CD$	158.6	0	13	13	13
A $\bar{B}\bar{C}D$	207.1	1	15	14	14
$\bar{A}BCD$	42.6	3	2	2	3
$\bar{A}BC\bar{D}$	17.3	0	4	5	4
$\bar{A}B\bar{C}D$	36.3	0	6	7	7
$\bar{A}B\bar{C}\bar{D}$	5.2	0	10	11	9
$\bar{A}\bar{B}CD$	13.9	0	9	9	11
$\bar{A}\bar{B}\bar{C}\bar{D}$	61.9	0	12	12	12
$\bar{A}\bar{B}CD$	171.7	0	14	15	15
$\bar{A}\bar{B}\bar{C}D$	429.7	1	16	16	16

关系数为: $\rho_{(I,II)} = 0.980$, $\rho_{(I,III)} = 0.988$, $\rho_{(II,III)} = 0.991$, 可以看出,模型 I、II 和 III 的后验概率高度相关,说明上述 3 个模型的后验概率空间分布形态非常相近. 事实上,根据后验概率圈定预测区的角

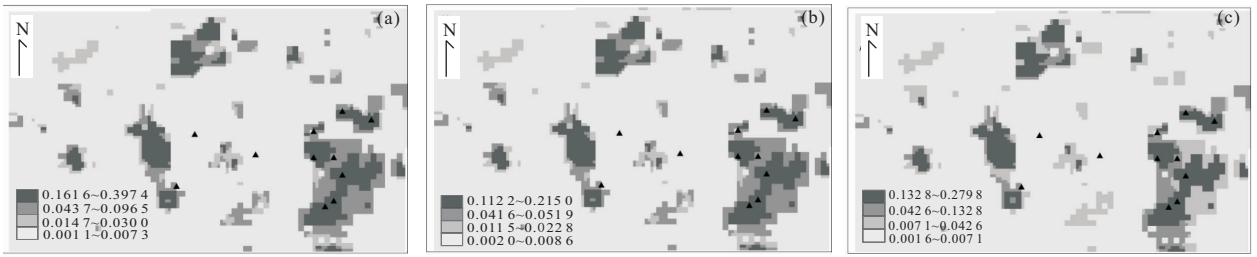


图 2 3 种证据权模型后验概率分级图

Fig. 2 Anomaly classification of posteriori probabilities obtained by three models

a. 模型 I 后验概率异常分级图; b. 模型 II 后验概率异常分级图; c. 模型 III 后验概率异常分级图. 黑色三角形表示 Sn 矿床

度,由表 2 可以看出,如果按照模型 III 的后验概率排序进行划分:1~3 划为一级异常区,占研究区总面积的 7.4%,包含了 6 个已知矿床,后验概率为 [0.1618, 0.2798];4~7 为二级异常区,占总面积的 6.8%,包含了 3 个已知矿床,后验概率为 [0.0425, 0.0929];8~11 为三级异常区,占总面积的 5.0%,包含了 0 个已知矿床,后验概率为 [0.0132, 0.0262];12~16 设为背景区,占总面积的 80.8%,包含 2 个已知矿床,后验概率为 [0.0016, 0.0070]. 后验概率异常分级如图 2c,从后验概率看,不同异常区及背景区内后验概率存在显著差异. 如果对模型 I 和 II 的后验概率按照相同的方法进行异常区分级,即,1~3 划为一级异常区,4~7 为二级异常区,8~11 为三级异常区,12~16 设为背景区(图 2a,2b). 由图 2 和表 2 看出,3 种模型的一级异常区是相同的,二级、三级异常区略有差别,背景区是相同的. 因此,从圈定预测区的角度,无论是证据权模型,还是改进的证据权模型圈定的预测异常区,差别都不是很大,这也进一步表明尽管证据图层不满足条件独立性假设,但用证据权模型圈定预测区仍然是一种有效的方法.

理论上 $\sum P(D | A_1 A_2 \dots A_n) \cdot P(A_1 A_2 \dots A_n) = \sum P(D | A_1 \dots A_n)$. 其中, $P(A_1 \dots A_n) = \frac{\text{area}(A_1 \dots A_n)}{\text{area}(S)}$,

$P(D) = \frac{n_D}{\text{area}(S)}$, 代入得到:

$$P(D | A_1 A_2 \dots A_n) \cdot \text{area}(A_1 A_2 \dots A_n) = n_D, \quad (15)$$

式中: $\text{area}(A_1 \dots A_n)$ 表示 $(A_1 \dots A_n)$ 的面积(单元数), $\text{area}(S)$ 表示研究区 S 的面积(单元数), n_D 表示训练点数. 公式(15)说明证据图层叠加后,所有不同子区域的面积与其相应后验概率的乘积之和应等于训练点数. 由于在矿产资源评价中,训练层一般是由已知矿床组成,训练点数就是已知矿床数. 因此,

公式(15)左边可以作为估计矿床数的方法,而矿床数估计是矿产资源评价的重要组成部分. 以此方法,用模型 I、II 和 III 的后验概率估计矿床数分别为 22.2、15.2 和 15.6,与 11 个已知矿床数相比,模型 II 和模型 III 比模型 I 有明显的改善.

5 结论

从模拟数据实验可知,证据变量不满足条件独立性导致普通证据权模型后验概率估计偏大或偏小,而加权证据权模型对于后验概率的估计有一定的改善;从应用实例看出,无论是逐步加权方法和无序加权方法,加权证据权模型对矿床数估计有明显改善. 加权证据权模型可以有效地减弱由于证据层不满足条件独立性假设对预测结果产生的影响.

模拟数据实验和应用实例结果表明,从后验概率从大到小排序的角度看,普通证据权模型、加权证据权模型,甚至后验概率理论值(尽管通常是未知的)具有极高的相关性,说明当证据权模型不满足条件独立性假设时,用证据权模型估计后验概率进行异常分级仍然是一种有效的方法. 这也正是证据权模型成为矿产资源评价中圈定预测靶区最广泛使用方法之一的原因.

理论上加权证据权模型的加权系数取值为 $(-\infty, +\infty)$,但本文用相关系数估计加权系数的方法,加权系数只能是 $[0, 1]$ 的值,如何将加权系数的估计值扩展到 $[0, 1]$ 之外,还需进一步探讨.

References

Agterberg, F., 2011. A modified weights-of-evidence method for regional mineral resource estimation. *Natural Resources Research*, 20 (2): 95 - 101. doi: 10. 1007/s11053-011-9138-0

Agterberg, F. P., 1989. Computer programs for mineral ex-

- ploration. *Science*, 245(4913): 76—81. doi: 10.1126/science.245.4913.76
- Agterberg, F. P., Bonham-Carter, G. F., Wright, D. F., 1989. Statistical pattern integration for mineral exploration. In: Gaál, G., Merriam, D. F., eds., *Computer applications in resource estimation prediction and assessment of metals and petroleum*. Pergamon Press, New York, 1—12.
- Agterberg, F. P., Cheng, Q. M., 2002. Conditional independence test for weights-of-evidence modeling. *Natural Resources Research*, 11(4): 249—255. doi: 10.1023/A:1021193827501
- Cheng, Q., 2012. Multiplicative cascade processes and information integration for predictive mapping. *Nonlinear Processes in Geophysics*, 19(1): 57—68. doi: 10.5194/npg-19-57-2012
- Cheng, Q. M., 2008. Non-linear theory and power-law models for information integration and mineral resources quantitative assessments. *Mathematical Geosciences*, 40(5): 503—532. doi: 10.1007/s11004-008-9172-6
- Cheng, Q. M., Agterberg, F. P., 1999. Fuzzy weights of evidence method and its application in mineral potential mapping. *Natural Resources Research*, 8(1): 27—35. doi: 10.1023/A:1021677510649
- Cheng, Q. M., Zhao, P. D., Chen, J. G., et al., 2009a. Application of singularity theory in prediction of tin and copper mineral deposits in Gejiu district, Yunnan, China; week information extraction and mixing information decomposition. *Earth Science—Journal of China University of Geosciences*, 34(2): 232—242 (in Chinese with English abstract).
- Cheng, Q. M., Zhao, P. D., Zhang, S. Y., et al., 2009b. Application of singularity theory in prediction of tin and copper mineral deposits in Gejiu district, Yunnan, China; information integration and delineation of mineral exploration targets. *Earth Science—Journal of China University of Geosciences*, 34(2): 243—252 (in Chinese with English abstract).
- Deng, M. F., 2009. A conditional dependence adjusted weights of evidence model. *Natural Resources Research*, 18(4): 259—258. doi: 10.1007/s11053-009-9101-5
- Journal, A. G., 2002. Combining knowledge from diverse sources; an alternative to traditional data independence hypotheses. *Mathematical Geology*, 34(5): 573—596. doi: 10.1023/A:1016047012594
- Krishnan, S., 2008. The tau model for data redundancy and information combination in earth sciences; theory and application. *Mathematical Geosciences*, 40(6): 705—727. doi: 10.1007/s11004-008-9165-5
- Krishnan, S., Boucher, A., Journal, A. G., 2005. Evaluating information redundancy through the tau model. *Quantitative Geology and Geostatistics*, 14(5): 1037—1046. doi: 10.1007/978-1-4020-3610-1_108
- Polyakova, E. I., Journal, A. G., 2007. The nu expression for probabilistic data integration. *Mathematical Geology*, 39(8): 715—733. doi: 10.1007/s11004-007-9117-5
- Schaeben, H., Boogaart, K. G., 2011. Comment on “a conditional dependence adjusted weights of evidence model” by Minfeng Deng in natural resources. research, 18(2009): 249—258. *Natural Resources Research*, 20(4): 401—406. doi: 10.1007/s11053-011-9146-0
- Zhang, S. Y., Cheng, Q. M., Zhang, S. P., et al., 2009. Weighted weights of evidence and stepwise weights of evidence and their application in Sn-Cu mineral potential mapping in Gejiu, Yunnan. *Earth Science—Journal of China University of Geosciences*, 34(2): 281—286 (in Chinese with English abstract).
- Zhang, S. Y., Wu, Q., Cheng, Q. M., et al., 2006. Weights of evidence method based on fuzzy training layer and its application in desertification assessment. *Earth Science—Journal of China University of Geosciences*, 31(3): 389—393 (in Chinese with English abstract).

附中文参考文献

- 成秋明, 赵鹏大, 陈建国, 等, 2009a. 奇异性理论在个旧锡铜矿产资源预测中的应用: 成矿弱信息提取和复合信息分解. *地球科学——中国地质大学学报*, 34(2): 232—242.
- 成秋明, 赵鹏大, 张生元, 等, 2009b. 奇异性理论在个旧锡铜矿产资源预测中的应用: 综合信息集成与靶区圈定. *地球科学——中国地质大学学报*, 34(2): 243—252.
- 张生元, 成秋明, 张素萍, 等, 2009. 加权证据权模型和逐步证据权模型及其在个旧锡铜矿产资源预测中的应用. *地球科学——中国地质大学学报*, 34(2): 281—286.
- 张生元, 武强, 成秋明, 等, 2006. 基于模糊预测对象的证据权方法及其在土地沙漠化评价中的应用. *地球科学——中国地质大学学报*, 31(3): 389—393.