

文章编号: 1006-544X (2007) 01-036-04

基于遗传算法的岩石地球化学元素关系评价

陈春香¹, 姜兰芳²

(1. 桂林工学院 电子与计算机系, 广西 桂林 541004; 2. 沈阳岩土工程技术测试开发中心, 沈阳 110015)

摘要:应用遗传算法的基因表达式程序,对山东齐家沟金矿的岩石地球化学数据进行了函数发现的实验研究.结果表明: Au-Ag、Au-As、Au-Sb 的 GEP 发现函数基本能反映训练样品元素之间的变化规律, Au-Ag、Au-As 为正相关的非线性函数关系, Au-Sb 为复杂的指数函数关系,且都能反映高异常值,说明遗传算法在地球化学元素评价中是有效的.

关键词:遗传算法; 基因表达式编程; 岩石地球化学; 元素评价

中图分类号: P628

文献标志码: A

在资源评价和找矿勘探中,岩石地球化学测量数据是非常重要的直接找矿信息:元素含量指示异常强度,元素关系指示原生的共生组合关系,可预测矿体类型、矿体埋深和成矿作用等^[1].岩石地球化学数据处理和评价方法关系到地球化学信息的有效显示和识别,进而影响着资源评价和找矿预测的成功率、客观性和有效性.国内外学者不断对岩石地球化学数据的处理和评价方法进行研究,提出如人工神经网络、分形等新方法,但由于地球化学数据分布型式以及数据结构的复杂性和多样性,使得一些方法的使用受限制.

基因表达式编程(Gene Expression Programming, GEP)^[2]遗传算法是一种新的智能算法,被广泛应用于数据挖掘和函数发现.本文采用遗传演化中的基因表达式编程算法,对岩石地球化学测量元素进行评价研究.数据选自山东齐家沟金矿的一个测区.

1 算法介绍

遗传算法是模拟生物在自然环境中的遗传和进化过程而形成的一种自适应全局优化的搜索算法,本质在于群体策略和遗传算子,群体搜索使遗传算法得以突破邻域搜索的限制实现整个解空间上的分布式信息搜索、采集和继承;遗传算子

利用适应值度量作为运算指标进行染色体的随机操作^[3].这样就使得遗传算法获得了强大的全局搜索能力,问题域的独立性、应用的鲁棒性等特征.

本次使用的是 Candida Ferreira 在 2001 年提出的基因表达式编程算法. GEP 的核心技术是基因表达式编码,采用固定长度的线性编码来表示个体(基因组),基因组由基因(Gene)通过函数集(Function Set)中的函数连接组成.该遗传算法结合传统遗传算法(GA)和遗传程序设计(GP)的优点,提出将个体先编码为固定长度的线性串,再表示成大小、形状都不同的非线性实体^[4]. GEP 是 GA 和 GP 的继承和发展,具有更强的解决问题的能力^[5].目前在数据挖掘和函数发现方面被广泛应用.在本项目研究中,对 Candida Ferreira 提出的 GEP 算法进行了一些改进,并用 VC 自行开发了适合函数发现的 GEP 程序.

在对岩石地球化学数据的 GEP 计算中,对参数设置进行了反复实验,表 1 是本次进行 GEP 计算的参数组合.对计算结果影响比较大的参数是适应度函数、函数集和演化代数.由于岩石地球化学数据变化大、比较离散,适应度函数只用相对值型函数,会出现高值无法拟合现象,而只用绝对值型函数,低值偏离较大,因此用复合型适

应度函数. 演化代数一般在3 000以后, 适应度函数值没有明显变化. 函数集选择+、-、×、/、ln()、exp()函数符, 增加其它函数符对计算没有明显影响. 其它参数的变化对计算影响不大, 用程序的默认值.

表1 GEP算法参数

Table 1 GEP parameter

参数名称	参数值
适应度函数	相对值类型0.05, 绝对值类型0.95
Evolution Strategy	$(u + \lambda)$ $\lambda = 10$
Selection Strategy	Roulette Wheel
Population Size	150
Generation	5000
函数集	+ - × / 指数函数 自然对数
基因参数	基因数=4; 基因头长=2; 基因身长=11; 同源基因头长=1; 同源基因身长=3
遗传操作算子	ISTransRate = 0.1; RISTransRate = 0.1;
选择概率	GeneTransRate = 0.1; 1 - pointRecRate = 0.3; 2 - pointRecRate = 0.3; GeneReRate = 0.1; MutationRate = 0.044

2 研究区地质概况

研究区位于胶东金矿集区齐家沟金矿东部, 在1.6 km²范围内开展1:5000的地球化学岩石测量, 其工作目的是: 查明测区内所测元素地球化学特征及富集规律, 圈出金矿化的富集部位, 为寻找外围金矿提供找矿靶区.

该区处在胶西北隆起区北部, 栖霞复背斜北翼, 燕山晚期郭家岭岩体东缘. 区内北东向断裂构造比较发育. 主要的断裂构造为大柳行-燕山河西断裂带. 在测区西南部以北, 发育2条规模不大的断裂构造, 分布有多条规模较大、走向北东20°的石英脉和闪长玢岩脉, 其余均为花岗岩, 属玲珑花岗岩体.

3 元素地球化学特征

表2是9种元素的地球化学特征值, 与胶东花岗岩平均值相比, 测区Au、Bi、Pb元素的平均含量明显高于花岗岩平均值, 且标准离差值较大, 这说明Au与Bi、Pb在本区的花岗岩中含量高, 离差大, 有相对富集的特征. Ag、As、Sb元素平均含量值与胶东地区的花岗岩和全省花岗岩的平均值相差不大, 但Ag离差较大, 说明在花岗岩中分布不均, 有一定的相对富集. Cu、Zn、Hg元素平均含量较低, 离差值较小, 说明本区Cu、Zn、

Hg元素相对富集较弱.

表2 研究区元素地球化学特征参数

Table 2 Geochemistry parameter of work area

元素	离差	最小值	最大值	测区 平均值	胶东花岗 岩平均值	全省 平均值
w(Au)/10 ⁻⁹	57.5	0.30	1514	6.78	1.08	1.56
w(Ag)/10 ⁻⁹	43.93	2.01	938.66	35.05	44.42	45.00
w(As)/10 ⁻⁶	8.51	0.90	225.40	3.03	3.64	6.14
w(Cu)/10 ⁻⁶	14.55	1.01	258.58	15.98	18.6	21.46
w(Pb)/10 ⁻⁶	20.60	2.44	510.36	37.90	25.6	21.06
w(Zn)/10 ⁻⁶	16.83	6.35	129.38	26.42	35.66	49.99
w(Sb)/10 ⁻⁶	0.86	0.02	23.00	0.34	0.21	0.46
w(Bi)/10 ⁻⁹	256.47	0.11	9169.56	97.38	50.0	0.21
w(Hg)/10 ⁻⁹	13.74	3.03	193.09	14.24	20.42	2.04

注: 参加统计的样品件数 $n = 1638$ (件). 数据来自文献[6]

选1号异常^[6]为GEP评价区, 该异常位于测区的东北角, Au异常呈条带状, 长×宽为300 m×50 m, 走向北东45°, 与构造走向一致, Au异常浓集中心明显, Ag、As等伴生元素出现异常, 是测区重点评价的异常.

4 元素关系评价

依据元素的地球化学性质, 元素在成矿中有一定的共生组合关系, 不同的元素组合代表不同的矿床类型. 元素关系研究能够评价和预测成矿类型, 所以确定异常元素的关系是岩石地球化学评价的一项重要工作^[7]. 传统的元素关系评价是计算元素之间的相关系数, 相关系数把元素的关系线性化, 只能简单的表示元素的大致关系, 实际观测表明, 任意2个元素之间的关系在不同的异常范围有不同的相关性, 即元素之间的关系是非线性的. 为了更准确表达元素之间的关系, 本文用GEP计算Au与Ag、As、Sb的关系表达式, 即关系函数.

4.1 Au-Ag关系评价

Au-Ag是地球化学性质很相近的金属成矿元素, 在胶东矿集区的大多数矿床中, Au、Ag有共生关系. 本异常区Au、Ag的含量有同步变化(图1之实线), 在Au含量低于 3×10^{-9} 样品中, Ag的含量变化小, 平均为 23.59×10^{-9} . 在Au含量 $(3 \sim 100) \times 10^{-9}$ 样品中(异常点), Ag的含量变化大, 平均为 46.02×10^{-9} , 最高为 173.92×10^{-9} . 在Au含量 $> 100 \times 10^{-9}$ 样品中(高异常点), Ag出现高含量, 平均为 245.16×10^{-9} , 最高为 507.04×10^{-9} .

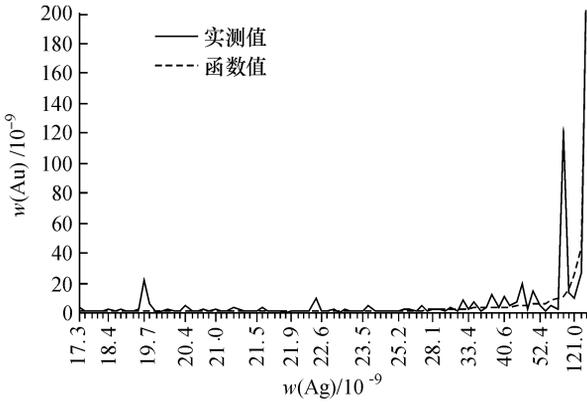


图1 Au-Ag 实测曲线与 GEP 曲线对比

Fig. 1 Au-Ag curve of measured value and GEP

对 1 号异常区样品的 Au、Ag 值进行 GEP 计算, 为了获得高适应值函数, 进行反复计算对比, 在表 1 的参数设置下, 适应度值在 390 ~ 430 之间变化. 图 1 虚线是适应度值为 426. 147 时的 GEP 函数曲线. 程序输出的 Au-Ag 基因表达式为:

$$0: = (\text{Exp}((3.94452)) - \text{Ln}((\text{Sqrt}(\text{Exp}((9.13938))))/(9.13938))))$$

$$1: = ((a - \text{Exp}(\text{Sqrt}(a))) + \text{Exp}(a))$$

$$2: = ((a - (9.34507)) * \text{Sqrt}(a))$$

$$3: = \text{Exp}(((4.75082) - (a + a)))$$

$$\text{同源基因} = (2/0)$$

0、1、2、3 分别代表 4 个基因, 其中的 a 表示参数 Ag, 同源基因选择了 0 与 2 号基因, 经化简, Au、Ag 的函数关系为

$$Y_{\text{Au}} = 0.02 X_{\text{Ag}}^{3/2} - 0.19 X_{\text{Ag}}^{1/2}$$

式中: Y 表示 Au 函数值; X 表示 Ag 自变量值.

发现的 Au-Ag 函数为二次函数, 当 Ag 在低值区 (<30) 变化时, Au 的含量值变化也小, 当 Ag 进入高值区 (>100 × 10⁻⁹) 时, Au 出现高值. 图 1 说明, Au-Ag 基因表达式函数基本反映了 Au、Ag 的关系变化关系, 所有训练样品的函数计算值与实测值的平均相对误差为 0.614.

4.2 Au-As 关系评价

As 是矿化剂元素, 是 Au 异常的重要示矿元素. 在本研究区, As 平均含量低于山东省平均含量, 是 As 的低背景区. 但其含量变化与 Au 含量变化相似, 在 Au 背景(含量低于 3 × 10⁻⁹)样品中, As 的含量低、变化小, 平均为 1.65 × 10⁻⁶. 在 Au 含量(3 ~ 100) × 10⁻⁶样品中(异常点), As 的含量变化大, 平均为 4.58 × 10⁻⁶. 在 Au 含量 > 100 × 10⁻⁶样品中(高异常点), As 出现高含量, 平均 50.96 × 10⁻⁶.

在表 1 参数条件下, 对 1 号异常区样品的 Au、As 值进行 GEP 计算, 适应值在 490 ~ 530 之间变化. 图 3 是适应值为 526. 585 时的实测值与 GEP 函数计算值的对比. 4 个基因如下:

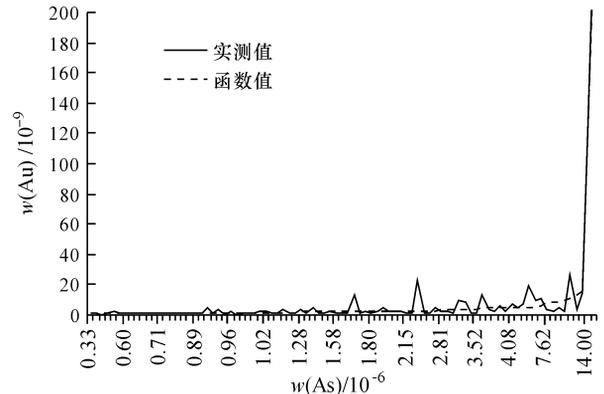


图2 Au-As 实测曲线与 GEP 曲线对比

Fig. 2 Au-As curve of measured value and GEP

$$0: = \text{Ln}((\text{Sqrt}(\text{Exp}(\text{Sqrt}((4.98398)))) + \text{Sqrt}(\text{Exp}((\text{Sqrt}(\text{Ln}((a + a))) * a))))))$$

$$1: = \text{Sqrt}(((a + (9.26695)) + (\text{Sqrt}(\text{Exp}((a/a))) * \text{Sqrt}(((a - a) * a))))))$$

$$2: = (\text{Ln}(\text{Sqrt}(\text{Exp}(\text{Sqrt}((\text{Sqrt}(\text{Sqrt}(\text{Exp}(a)))) * ((a/a)/\text{Sqrt}(a)))))) + a)$$

$$3: = \text{Sqrt}(\text{Ln}(\text{Ln}((\text{Exp}((\text{Ln}((3.91827) + a))) * ((a + a) * \text{Sqrt}(a)))) - \text{Ln}(a))))$$

$$\text{同源基因} = \text{Ln}(\text{Exp}(2))$$

0、1、2、3 分别代表 4 个基因, 其中的 a 表示参数 As, 同源基因选择了 2 号基因, 经化简, Au、As 的函数关系为

$$Y_{\text{Au}} = X_{\text{As}} + \frac{1}{2}((e^{X_{\text{As}}})^{1/4} X_{\text{As}}^{-1/2})^{1/4}$$

式中: Y 表示 Au 函数值; X 表示 As 自变量值.

发现的 Au-As 函数是一个比较简单指数函数, 函数中的指数部分随 As 的指数. 图 2 说明, Au-As 基因表达式函数基本反映了 Au、As 的关系. 所有样品的函数计算值与实测值的平均相对误差 0.717.

4.3 Au-Sb 关系评价

在胶东矿集区, Sb 没有被列入 Au 矿的元素组合^[7]. 本研究区, 在 Au 低异常样品中(含量(3 ~ 100) × 10⁻⁹), Sb 的含量变化很大, 没有变化规律. 但在 Au 高异常样品中(含量 > 100 × 10⁻⁹), Sb 出现高含量(图 3), 最高 1 720.00 × 10⁻⁶.

