

遗传程序设计在多金属 成矿预测中的应用^①

蔡煜东

(中国科学院上海冶金研究所)

提 要 本文运用遗传程序设计,对鄂东南地区 44 个铁帽进行了计算机识别,识别成功率较高。结果表明,该方法性能良好,可望成为多金属成矿预测的一种有效的辅助手段。

关键词 多金属成矿预测 遗传程序设计

1 遗传程序设计

1975 年,Holland^[1]受生物学中“生物进化”和“自然选择”学说的启发,提出了著名的遗传算法^[1](Genetic Algorithm)。经过近 20 年的研究、应用,遗传算法已成为非线性优化和系统辨识的有效工具,被广泛应用于机器人系统及神经网络学习过程,以解决 NP 完全性、规划控制等问题,取得了很好的效果^[2,3]。

90 年代初,Koza 在以数量(整数、实数)为对象的遗传算法的基础上,提出了以计算机程序为对象的遗传程序设计^[4](Genetic Programming)。遗传程序设计是一种最优化方法,即通过对问题进行程序结构化处理(一般为树结构表示),称为“基因”,给出一种进化函数(或称压力选择函数),通过某些遗传运算,如交叉重组、突变等,将那些“最合适的”保留下来。

近几年来,随着遗传程序设计的理论及应用的不断进步,该方法被逐渐应用于模式识别、优化控制、决策支持系统等领域,取得了很好的成果^[4]。

遗传程序设计的流程图如图 1 所示:

① 收稿日期 1994. 10. 5 改回日期 1995. 1. 4

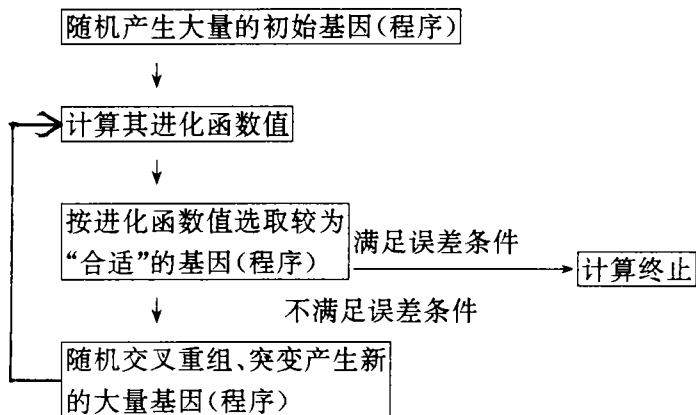


图1 遗传程序设计流程图

Fig. 1 Genetic programming frame

运用遗传程序设计处理一个具体问题包括如下4个方面:

- (1)问题的程序结构化处理(即将问题进行树结构表示转化为“基因”)。
- (2)进化函数的选择。
- (3)遗传程序设计中的运算方式及运算效率的选择(交叉重组(率)、突变(率)等)。
- (4)遗传程序设计中的参数选择(基因数量、遗传世代数等)。

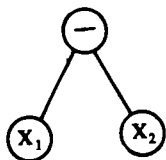
有关遗传程序设计的详细论述请参看文献^[4]。

下面具体阐述遗传程序设计的上述4个方面:

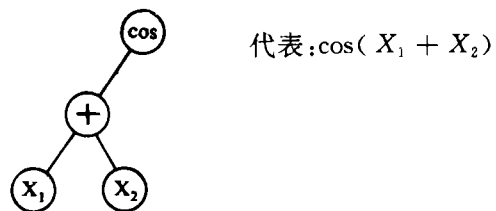
例:对于样本集 $(X_1, X_2) \rightarrow Y_1$, 建立模拟函数

1)将该问题进行结构化处理,即用“树结构”表示模拟函数,转化为“基因”:

如:



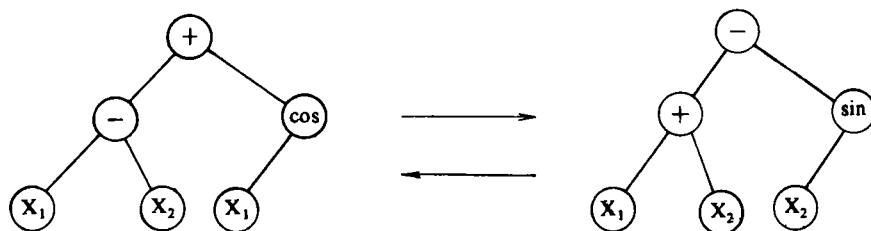
代表 $X_1 - X_2$;



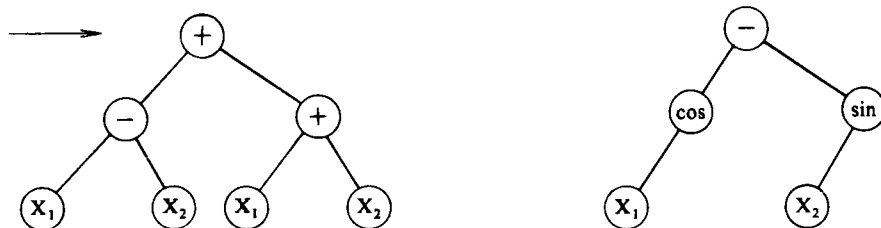
(2) 进化函数选择 $\min(\bar{Y}_1 - Y_1)^2$, 即模拟函数值与真实函数值之间的误差为最小。

(3) 遗传运算方式选择交叉重组、突变两种遗传运算:

(a) 交叉重组:

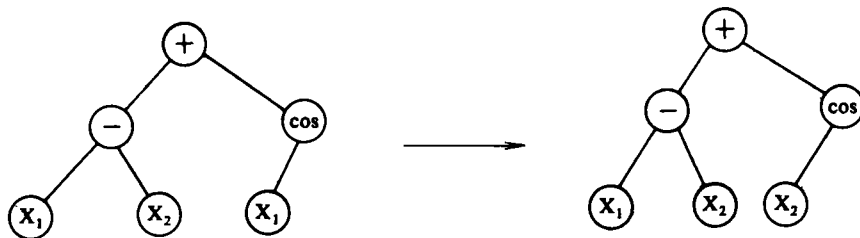


上一代基因



新一代基因

(b)突变:



(4)算法中的参数——基因数量,即多少基因参与遗传运算,一般选取 20~40;遗传世代,即遗传多少代终止,一般选取 5~20。

2 遗传程序设计应用于铁帽识别

2.1 资料来源

文献^[6]报道了鄂东南地区若干个铁帽(包括铁矿和铜矿)的 8 种化学元素和成矿的有关数据。本文以此作为研究对象,对鄂东南地区 44 个铁帽进行计算机识别,尝试了遗传程序设计方法的效果。

2.2 识别模型的建立

按照遗传程序设计,笔者用 C 语言编制了铁帽识别程序,并在 AST/PPA 微机上通过运行。

首先,随机选取其中 36 个样本(第 1 类:铁矿;第 2 类:铜矿)作为拟合数据,以其特征变量(Cu、Pb、Zn、Ag、Mo、W、Bi、Sn)作为输入,两类样本期望的输出依次定为:“正”(第 1 类);“负”(第 2 类)。遗传程序设计的运算方式及效率、算法参数的选取见表 1。经过“遗传进化”,所得结果程序能正确地划分这些样本,建立了化学元素含量与成矿专属性之间的复杂对应关系,见表 2 和图 2。

2.3 识别成功率

为了考验所建立的模型,将未参加建模的 7 个样本,作为“未知”样本,已掌握了知识信息的模型(结果程序)对其进行识别。按其输出决定其类别。由表 3 可见,识别结果与实际情况完全一致。

表 1 遗传程序设计中的进化函数、运算方式、效率及参数的选取

Table 1 Selection of evolution function operation mode,efficiency and parameter in genetic programming

交叉重组率(C)	突变率(M)	基因数量(P)	遗传世代数(G)	进化函数(EF)
0.2—1.0(步长:0.05)	0.1	50	5	判对样本数

表 2 36 个建模样本

Table 2 36 specimen for model establishment

Ca	Pb	Zn	Ag	Mo	W	Bi	Si	期望输出	实际输出	类别
0.11	0.08	0.13	0.05	0.17	0.13	0.3	0.3	正	0.435348	1
0.03	0.03	0.06	0.03	0.09	0.1	0.3	0.3	正	0.440651	1
0.11	0.05	0.09	0.03	0.09	0.1	0.25	0.3	正	0.409979	1
0.05	0.14	0.32	0.09	0.08	0.1	0.25	0.3	正	0.507152	1
0.3	0.11	0.28	0.18	0.2	0.1	0.25	0.3	正	0.312431	1
0.42	0.01	0.06	0.08	0.39	0.28	0.25	0.9	正	0.151334	1
0.75	0.03	0.07	0.05	0.04	0.1	0.25	0.4	正	-0.588064	1
0.01	0.01	0.14	0.03	0.08	0.1	0.25	0.3	正	0.422135	1
0.23	0.01	0.26	0.03	0.04	0.1	0.25	0.3	正	0.314633	1
0.44	0.07	0.14	0.06	0.17	0.11	0.3	0.3	正	0.141074	1
0.19	0.07	0.12	0.02	0.08	0.1	0.25	0.3	正	0.377843	1
0.02	0.01	0.08	0.03	0.2	0.12	0.25	0.3	正	0.422602	1
0.21	0.02	0.21	0.03	0.05	0.14	0.25	0.3	正	0.337119	1
1.0	0.03	0.1	0.1	0.03	0.1	0.25	0.3	负	-0.828351	2
0.4	0.02	0.01	0.03	0.04	0.16	0.25	0.4	负	0.174367	2
1.0	0.06	0.3	0.2	0.22	0.15	0.25	0.3	负	-0.701470	2
1.0	0.66	0.24	0.82	1.0	0.13	0.25	0.3	负	-0.269267	2
1.0	0.16	0.21	0.32	1.0	0.33	0.3	0.8	负	-0.577703	2
1.0	0.03	0.28	0.36	0.52	0.22	0.3	0.8	负	-0.716154	2
0.07	0.17	1.0	0.09	0.07	0.1	0.3	0.3	负	0.520783	2
0.25	1.0	1.0	1.0	0.38	0.25	0.3	0.5	负	0.105774	2
0.67	0.04	0.07	0.18	0.04	0.11	0.35	0.4	负	-0.312425	2
1.0	0.03	0.15	0.65	1.0	1.0	0.45	1.0	负	-1.140495	2
1.0	0.01	0.31	1.0	0.05	0.1	0.7	0.5	负	-0.758934	2
0.81	0.43	0.53	0.16	0.11	0.11	0.3	0.3	负	-0.273977	2
0.07	1.0	1.0	1.0	0.32	0.11	0.25	0.4	负	-0.013301	2

表3 7个待判样本
Table 3 7 specimen for model establishment

Cu	Pb	Zn	Ag	Mo	W	Bi	Sn	计算输出	类别(预测)	类别(实际)
0.16	1.0	0.3	1.0	0.13	0.21	1.0	0.4	0.200549	1	1
0.16	0.14	0.12	0.13	0.17	0.11	0.25	0.3	0.443710	1	1
0.02	0.04	0.09	0.03	0.05	0.1	0.25	0.3	0.442610	1	1
0.1	0.02	0.11	0.03	0.31	0.1	0.25	0.3	0.390856	1	1
1.0	0.1	0.38	0.2	0.44	1.0	1.0	1.0	-0.554567	2	2
1.0	0.05	0.25	1.0	0.38	0.25	0.6	1.0	-0.696643	2	2
0.45	1.0	1.0	0.87	0.6	0.13	1.0	0.5	-0.022916	2	2

3 讨论与结论

1. 遗传程序设计用于铁帽识别,是在程序空间搜索一最佳程序(非线性模型),使之能较好地划分有关类别,即可在有不完备知识的基础上,寻找正确的物理模型。
2. 运用遗传程序设计研究铁帽识别,运算方式及效率、参数的选择是个关键,对算法参数(基因数量、遗传世代数)可尽量范围选择大些,范围大小只影响搜索时间的长短,而编成程序后,计算机实现拟合十分简便。突变率一般不宜过大(0.05~0.15),防止变异过大丧失遗传中保留的较为“合适的基因”。交叉重组率应是动态变化,由小至大。
3. 遗传程序设计不仅可用于铁帽识别,对于一般的非线性建模,只要在某个目标函数(对应于算法中的进化函数)下,都可进行模型(结构)的遗传运算,得出最佳的非线性模型。因此,遗传程序设计对各类地质体识别具有普遍意义。

参考文献

1 Goldberg D. Genetic Algorithm in search,optimazation,and machine learning,addison—wesley,reading,MA. ,1989

2 Davidor Y. Genetic algorithm and robotics,A heuristic strategy for optimization. World scientific pub. Co. ,New jersey, USA,1991

3 Belew R K. McInemey J. and Schraudolph N N. “Evolving networks;using the genetic algorithm connectionist learning,” Artificial life II,Ed. Langton C G,etc. ,Addison—wesley,1991,pp. 511—547

4 Koza. J. R Genetic Programming,MIT Press,Cambridge,Mass. ,1992.

5 刘文斌,等. 物探化探计算技术,1991,13(2):105—109.

APPLICATION OF GENETIC PROGRAMMING ON PREDICTION OF MULTIMETALLIC DEPOSITES

Cai Yudong

(Shanghai Institute of Metallurgy, Chinese Academy of Sciences, Shanghai, 200050)

Abstract

Genetic programming was applied to the prediction of the gossan in the southeast of Hubei province. The successful rate was high. The results showed that the performance of genetic programming was good and therefore it might be referred as an effective assistant technique for prediction of multimetallic deposits.