

地震序列类型预测的遗传程序设计方法*

蔡煜东

(中国上海 200050 中国科学院上海冶金研究所)

摘要 提出预测地震序列类型的遗传程序设计方法,并选取一组实例作为研究对象,验证了该方法的可靠性。结果表明,遗传程序设计方法性能良好,可望成为地震序列类型预测的有效工具。

主题词 地震序列 预测 遗传程序设计

引言

地震是以不同序列的形式出现的。一次地震发生后,紧接着是否还有更大的地震,这是地震预报中一个重要的实际问题,也就是预测地震序列的类型问题。

本文运用近几年发展起来的遗传程序设计方法,进行地震序列类型的预测研究。有关这方面的研究工作尚未见报道。

1 遗传程序设计

1975年, Holland 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(genetic algorithm)(Goldberg, 1989)。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(Davidor, 1991; Belew *et al.*, 1991)。

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(genetic programming)(Koza, 1992)。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。

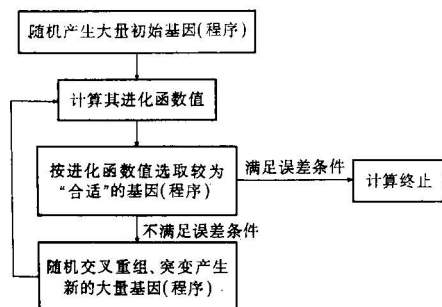


图1 遗传程序设计流程图

* 1994-10-05 日收到初稿, 1994-11-28 日收到修改稿。

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(Koza, 1992)。

遗传程序设计的流程图如图 1 所示。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)。

下面具体阐述遗传程序设计的上述 4 个方面:

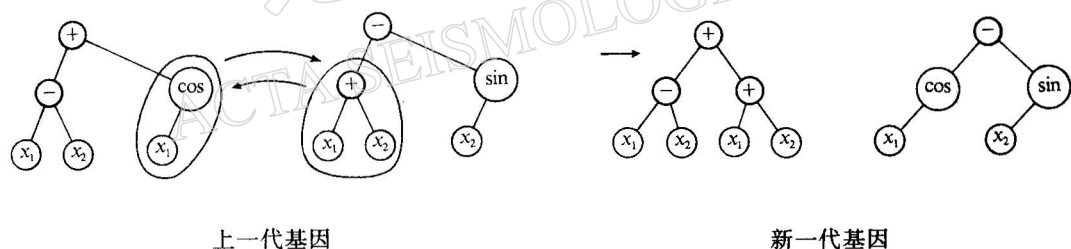
例 对于样本集 $(x_1, x_2) \rightarrow y_1$, 建立模拟函数。

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”。如下图所示。

2) 进化函数选择 $\min(\hat{y}_1 - y_2)^2$, 即模拟函数值与真实值之间的误差为最小。

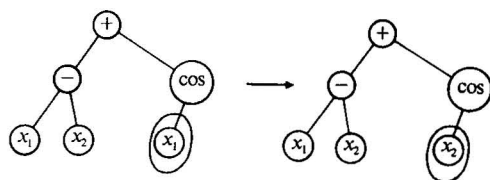
3) 遗传运算方式选择交叉重组、突变两种遗传运算:

a) 交叉重组



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预测地震序列类型

按不同地震序列的特点,通常可把它们分为 3 类:① 孤立型或单发式地震型。前、余震所占的能量百分比都非常小;② 多主震型或震群型,在整个序列中有两个或两个以上震级相近的较大地震(主震);③ 主震余震型。有一个震级较大的主震,但余震较多,且占有较大一些的能量比。

在冯德益等(1982)的基础上,我们选用以下 3 项指标作为判断震型的基本指标。这 3 项指标均可在第一个强震发生后不久得到,即:

- (1) $\Delta M_{\text{前}}$, 序列中第一个强震 M_1 与其最大前震的震级之差;
- (2) E_2/E_1 , 第一个强震发生后两天内余震释放的能量与第一个强震能量之比;

(3) E_5/E_0 . E_5 为第一个强震发生后 5 天内释放的地震能量(包括 E_1), 与根据余震体积分估计出的地震总能量之比.

为简化问题, 我们把孤立型和主震余震型合并为一类, 统称单主震型, 并要求预测单主震型和多主震型这两大类地震序列. 试用遗传程序设计解决这一问题.

本文选取 15 例地震序列类型作为研究对象, 见表 1.

表 1 原始数据

地 震	M	$\Delta M_{前}$	E_2/E_1	E_2/E_0	类 型
通海	7.7	3.5	0.33%	1.4%	单主震型
炉霍	7.9	6.9	0.31%	2.4%	单主震型
渤海	7.4	4.1	0.11%	1.1%	单主震型
海城	7.3	2.7	0.45%	1.3%	单主震型
永善	7.1	3.3	0.07%	2.4%	单主震型
柯坪	6.2	5.2	0.30%	0.9%	单主震型
马边	6.0	1.3	91.8%	0.6%	多主震型
康定	5.7	4.2	0.01%	1.4%	多主震型
松潘	7.2	6.2	0.28%	0.6%	多主震型
邢台	7.2	1.6	1.33%	0.3%	多主震型
龙陵	7.4	2.1	149%	2.0%	多主震型
唐山	7.8	6.2	14.7%	1.6%	单主震型
甘孜	6.8	2.6	0.63%	1.0%	单主震型
河间	6.3	4.8	0.05%	0.9%	单主震型
东川	6.8	3.3	1.37%	0.6%	多主震型

按照遗传程序设计, 笔者用 C 语言编制了地震序列类型预测程序, 并在 AST/PP4 微机上通过运行.

随机选取 13 例作为拟合数据, 以其特征变量($\Delta M_{前}$, E_2/E_1 , E_5/E_0)作为输入, 将两类地震序列(两类样本)的期望输出依次定为“正”(单主震型), “负”(多主震型). 遗传程序设计的运算方式及效率和算法参数的选取见表 2. 经过“遗传进化”, 所得结果程序能较正确地识别这些样本, 建立了特征变量与序列类型之间的复杂对应关系, 见表 3 和图 2.

表 2 遗传程序设计中的进化函数、运算方式、效率及参数的选取

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 3 13 个建模样本

地 震	M	$\Delta M_{前}$	E_2/E_1	E_2/E_0	期望输出	实际输出	类别
通海	7.7	3.5	0.33%	1.4%	正	0.443138	1
炉霍	7.9	6.9	0.31%	2.4%	正	1.119091	1
渤海	7.4	4.1	0.11%	1.1%	正	0.221948	1
唐山	7.8	6.2	14.7%	1.6%	正	0.435857	1
甘孜	6.8	2.6	0.63%	1.0%	正	-0.189962	1
河间	6.3	4.8	0.05%	0.9%	正	0.118296	1
海城	7.3	2.7	0.45%	1.3%	正	0.252391	1
永善	7.1	3.3	0.07%	2.4%	正	0.881595	1
康定	5.7	4.2	0.01%	1.4%	负	0.620223	2
松潘	7.2	6.2	0.28%	0.6%	负	-0.048796	2
邢台	7.2	1.6	1.33%	0.3%	负	-0.527147	2
龙陵	7.4	2.1	149%	2.0%	负	-0.121059	2
东川	6.8	3.3	1.37%	0.6%	负	-0.334109	2

为了考验所建立的模型(结果程序), 将未参加建模的 3 个地震序列, 作为“未知”样本, 由已掌握了知识信息的神经网络对其进行预测. 由表 4 可见, 预测结果与实际情况全部一致.

3 讨论和结论

(1) 遗传程序设计用于地震序列类型预测, 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, 既可在无任何先验知识的情形下, 建立单纯的统计模型, 亦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地震序列类型预测, 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)其范围可尽量选择大些. 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, 应由小至大.

(3) 遗传程序设计不仅用于地震序列类型预测的建模, 对于一般的非线性建模, 只要

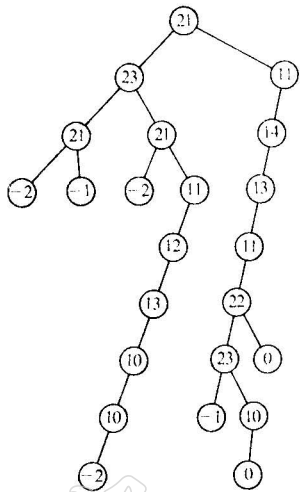


图 2 树结构表示(20=“+”, 21=“-”, 22=“*”, 23=“./(1.1+1)”, 10= x^3 , 11= $\sin x$, 12= $\operatorname{tg} x$, 13= $(e^x - e^{-x})/2$, 14= x ; $i=0, -1, -2, \dots, -n$ 表示第 i 个输入变量)

表 4 两个预报样本

地震	M	$\Delta M_{\text{前}}$	E_2/E_1	E_2/E_0	计算输出	类别(预测)	类别(实际)
柯坪	6.2	5.2	0.30%	0.9%	0.132428	1	1
马边	6.0	1.3	91.8%	0.6%	-0.376271	2	2

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, 遗传程序设计对地震研究中各类非线性建模、预测问题具有普遍意义.

天津市地震局冯德益研究员提供了全部地震序列的有关数据, 特此致谢.

参 考 文 献

冯德益, 林命遇, 顾瑾平, 1982. 模糊综合评判与贴近度在震源研究中的应用. 地震, (3): 29~45
Belew R K, McInemey J, Schraudolph N N, 1991. *Artificial Life I*. Addison-Wesley, MA. 511~547
Davidor Y, 1991. *Genetic Algorithm and Robotics. A Heuristic Strategy for Optimazation*. World Scientific Pub. Co., New Jersey, USA. 431~438
Goldberg D, 1989. *Genetic Algorithm in Search, Optimazation, and Machine Learning*. Addison-Wesley, Reading, MA. 123~128
Koza J R, 1992. *Genetic Programming*. MIT Press, Cambridge, Mass. 41~45