

遗传程序设计用于工业调优 ——炼钢转炉炉龄预测¹⁾

蔡煜东 程兆年 陈念贻 杨传仁 许 立

(中国科学院上海冶金研究所 上海 200050)

陆文聪

(上海大学化学系 上海 201800)

摘 要 提出了炼钢转炉炉龄问题预测的遗传程序设计方法,并选取一组标样进行了研究.结果表明,该方法性能良好,可望成为钢铁工业预测、调优的一种有效辅助手段.

关键词 工业调优,炼钢转炉炉龄,遗传程序设计.

1 前言

化学模式识别方法用于工业调优已有多多年,并取得了一定的经济效益^[1].

我们运用近几年发展起来的遗传程序设计这一新颖的化学模式识别方法,建立了炼钢转炉炉龄预测和调优的计算机智能专家系统,以期找到某些规律来提高转炉寿命.

2 遗传程序设计

1975年,Holland^[2]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^[2](genetic algorithm).经过近20年的研究和应用,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^[3,4].

九十年代初,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^[5](genetic programming).遗传程序设计是一种对问题进行程序结构化处理(一般为树结构,称为“基因”)的最优化方法.其目的是构造一个能表达输入参数与输出参数间复杂对应关系的最佳“基因”,而“基因”的特定树状结构决定了一种特定的运算过程.寻找最佳“基因”的过程是一种“适者生存”的进化过程.因此,在遗传程序设计过程中,需定义一种进化函数,进化函数的具体形式取决于研究对象的某个特性,比如本工作中用“基因”对样本集中样本所属类别进行判别,并以判对样本数作为进化

¹⁾ 国家科委基础研究项目资助课题.

函数的取值. 进化函数的值将作为选取最佳“基因”的判据. 遗传运算包括交叉重组、突变等, 即改变“基因”的排列方式和运算方式, 而“最合适的基因”对应于最佳的进化函数值. 最终将“最合适的”基因保留下来.

近几年来, 随着遗传程序设计的理论及应用的不断发展, 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.

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 (即将问题转化为以树结构表示的“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, 即遗传程序设计的循环次数等).

有关遗传程序设计的详细论述请参看文献 [5].

3 遗传程序设计用于炼钢转炉炉龄预测

某钢铁厂炼钢用转炉的炉龄目前为每个炉役约为 1000 炉钢. 按目前 30 炉/天的炼钢规模, 大约一个月左右就需停炉一次, 替换炉衬. 每次替换炉衬的花费需数百万元. 为了挖掘生产潜力, 提高经济效益, 该厂提出要力争在几年内使炉龄得到大幅度的提高. 影响炉龄问题的因素很复杂, 我们仅就所提供的炉役平均数据, 用遗传程序设计方法进行处理, 以期找到解决问题的方向.

厂方提供的数据含 33 个炉役, 将每个炉役视为一个样本. 每个样本含 18 个变量. 每个变量都是该炉役中各炉的平均值. 样本集见表 1.

表 1 25 个建模样本

x_1	x_3	x_5	x_{18}	期望输出	实际输出	类别
0.2436	17.7	37.2	37.9	正	1.279094	1
0.2804	18.3	37.5	20.3	正	0.072114	1
0.197	17.3	35.9	57.4	正	0.724487	1
0.1524	17.6	36.0	36.7	正	0.885221	1
0.2922	18.5	41.4	83.3	正	1.376447	1
0.2244	18.8	37.7	49.0	正	0.451072	1
0.2155	20.2	40.2	41.7	正	-1.048746	1

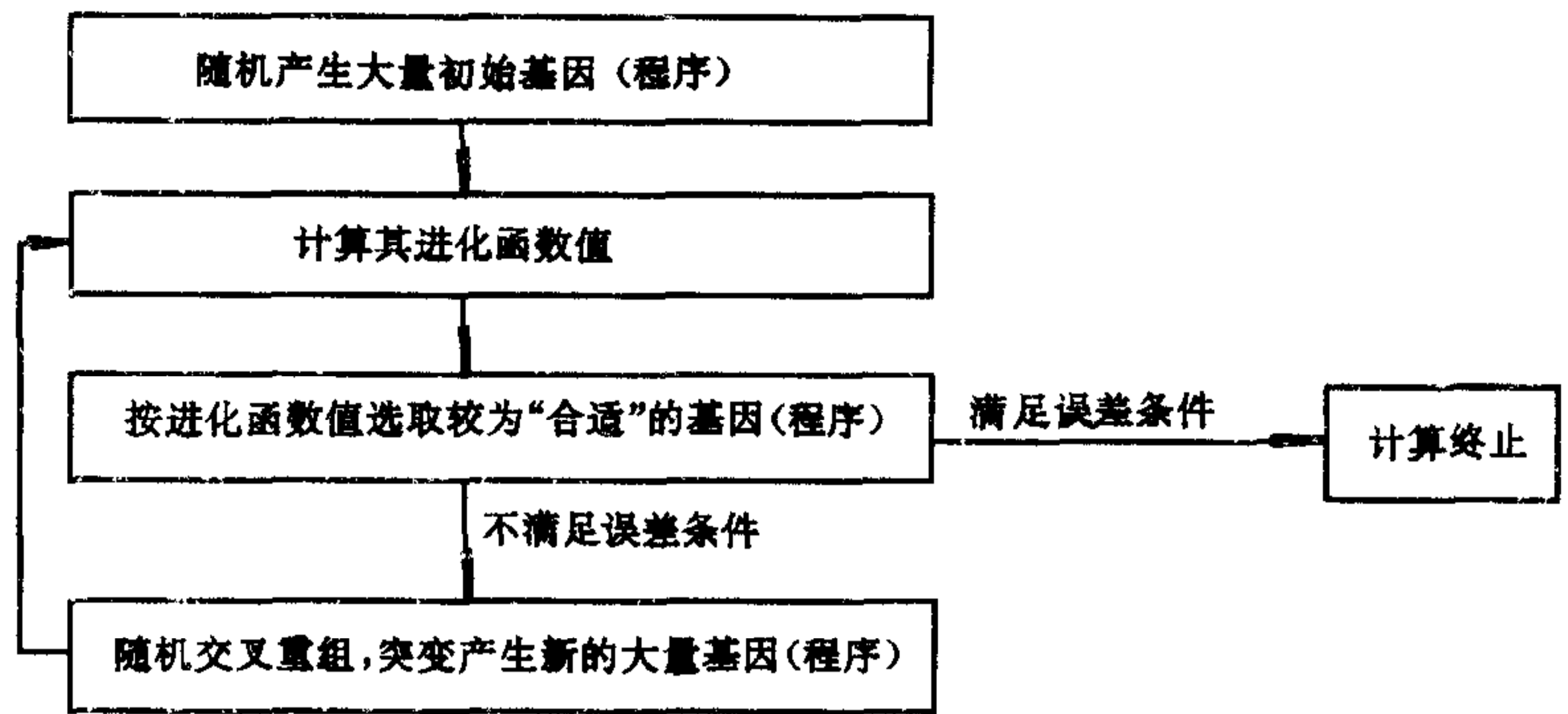


图 1 遗传程序设计流程图

续表

x_1	x_3	x_5	x_{18}	期望输出	实际输出	类别
0.2685	17.7	38.6	78.9	正	0.554854	1
0.2348	18.0	39.4	57.4	正	1.203874	1
0.418	18.8	41.0	26.7	正	1.412103	1
0.184	16.2	35.3	44.8	正	1.214598	1
0.1963	18.1	37.2	63.0	正	0.774568	1
0.1353	19.0	38.6	39.7	正	-1.335123	1
0.1835	18.9	41.8	47.2	负	-1.197385	2
0.1386	18.9	40.5	22.5	负	-1.550489	2
0.2083	18.3	39.8	52.6	负	0.740887	2
0.103	18.4	39.2	35.0	负	-1.464388	2
0.4893	19.3	41.4	31.3	负	1.478446	2
0.1854	19.0	40.8	28.6	负	-1.332643	2
0.23	18.4	37.1	21.6	负	-0.485523	2
0.2058	19.0	40.0	41.2	负	-0.774732	2
0.0925	17.9	38.7	66.7	负	-0.522403	2
0.0316	20.9	41.2	52.6	负	-3.195589	2
0.1465	19.1	38.6	28.1	负	-1.444614	2
0.1008	18.2	37.0	33.9	负	-0.942572	2

样本中的 18 个变量如下:

x_1 : 喷补料量(kg/吨钢)

x_2 : 一次命中率(%)

x_3 : 吹炼时间

x_4 : 等待时间

x_5 : 炼钢时间

x_6 : 铁水含 Si 量

x_7 : 铁水含 Mn 量

x_8 : 铁水含 P 量

x_9 : 铁水含 S 量

x_{10} : 铁水温度

x_{11} : 终点温度

x_{12} : 石灰加入量(kg/吨钢)

x_{13} : 白云石加入量(kg/吨钢)

x_{14} : 铁水比(%)

x_{15} : 渣中总铁量

x_{16} : 渣中 MgO(%)

x_{17} : 渣的碱度(CaO/SiO₂)

x_{18} : 作业率(%)

数据处理时将样本分为二类. 1 类是好类样本, 指炉龄大于 1000 炉者; 2 类是差类样本, 指炉龄小于 1000 炉者. 1 类样本有 16 个, 2 类样本 17 个.

首先, 运用一般的模式识别要因分析技术^[6], 对上述 18 个变量进行要因分析. 经过筛选, 发现 x_5 (炼钢时间)、 x_3 (吹炼时间)、 x_1 (喷补料量)和 x_{18} (作业率)是决定炉龄的主要因素. 上述影响转炉炉龄 18 个因素的模式识别要因分析已经报道^[7].

按照遗传程序设计, 笔者用 C 语言编制了炉龄预测程序, 并在 AST/PP4 微机上通过运行.

选取其中 25 个样本作为遗传程序设计的拟合数据, 以其特征变量(x_5 、 x_3 、 x_1 、 x_{18})作为输入, 二类样本的期望输出分别变为“正”(1 类)、“负”(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 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(x_5 、 x_3 、 x_1 、 x_{18})与炉龄之间的复杂对应关系, 见表 1 和图 2.

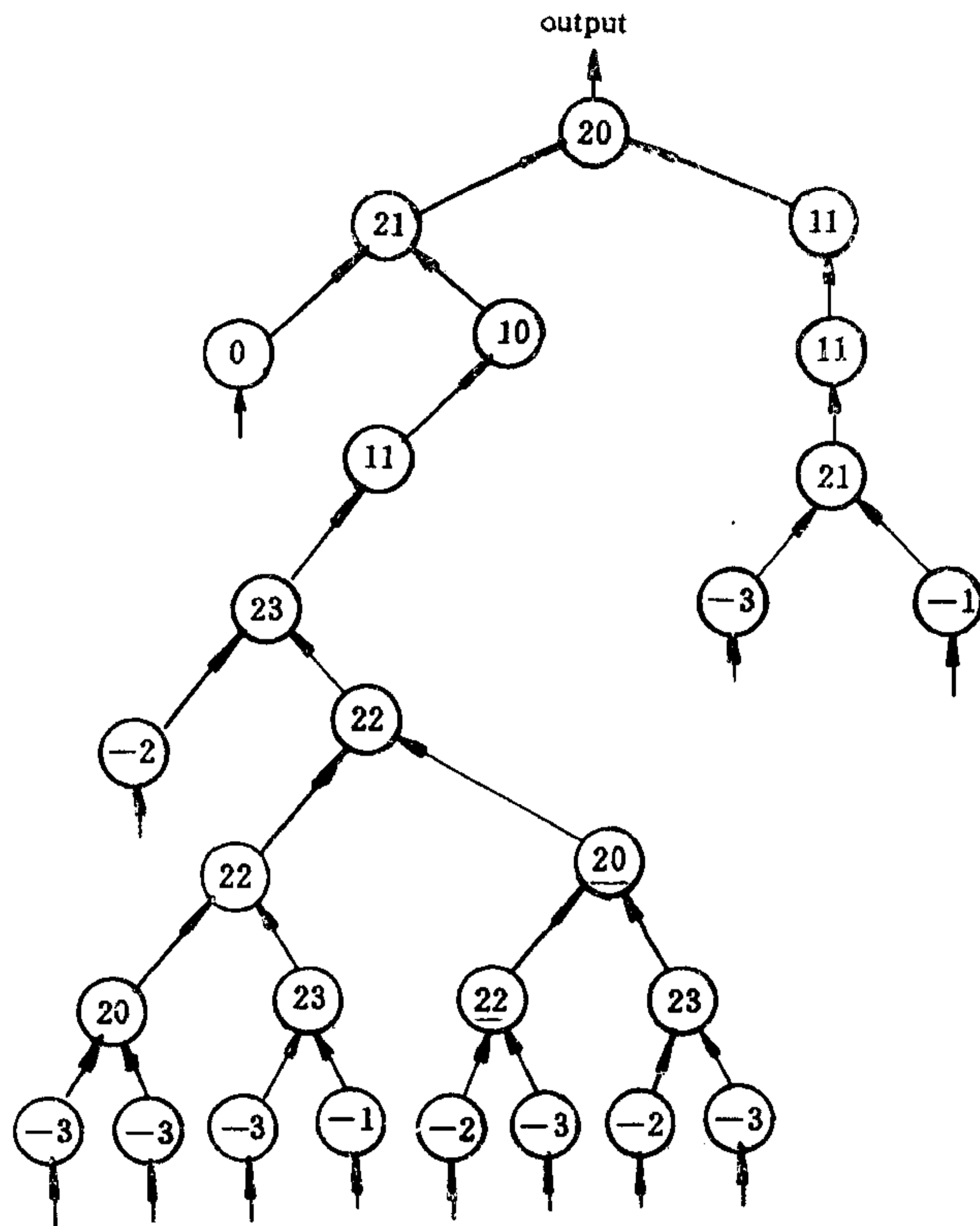


图 2 树结构表示(基因优化结果)

20=“+”,21=“-”,22=“*”,23=“*/(|*|+1)”,10= x^3 ,11= $\sin x$,12= $\operatorname{tg} x$,
 13= $(e^x - e^{-x})/2$,14= x , $i=0, -1, -2, \dots, -n, \dots$,表示一个输入样本的第 i 个特征变量

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

图 2 中树结构的每一圆圈表示一个节点,箭头表示输入输出的方向,圆圈内的代码表示一种运算. 树结构各末端用来输入样本的特征变量,顶端用来输出样本的特性. 一位数字的代码 0, -1, -2, -3 分别表示输入样本的特征变量 x_1, x_3, x_5, x_{18} 值,而二位数字的代码表示一种运算. 从输入样本的特征变量至输出特性的过程,由树结构中间的那些特定的运算所决定. 比如“ $10 = x^3$ ”表示将代码为 10 的节点前面的一个节点的输出值 C 堆积 C^3 运算后的值作为 10 节点的输出值. 又比如 $23 = “*/(|*|+1)”$ 表示将代码为 23 的节点前的两个节点的输出值 a 和 b 作“ $a * b / (|a * b| + 1)$ ”运算后的值作为 23 节点的输出值. 其它类推. 表 1 中列出的“实际输出”就是图 2 中树结构顶端的输出值.

为了进一步考察所建立的模型,将未参加建模的 8 个样本作为“未知”样本,由已掌握了知识信息的模型(结果基因程序)对其进行预测,按其输出决定其类别. 由表 3 可见,预测结果与实际情况完全一致.

表 3 8 个预测样本

x_1	x_3	x_5	x_{18}	计算输出	类别(预测)	类别(实际)
0.2436	17.7	37.2	37.9	1.279094	1	1
0.1679	17.1	34.6	37.3	1.247005	1	1
0.2222	18.4	37.0	42.9	0.804085	1	1
0.2445	18.2	36.6	37.9	1.101652	1	1
0.2702	18.9	39.5	31.3	-0.104267	2	2
0.0491	20.3	40.6	35.0	-2.877597	2	2
0.1487	19.4	39.5	33.3	-1.507794	2	2
0.1298	18.4	37.2	44.3	-0.253483	2	2

4 讨论和结论

1) 遗传程序设计用于炉龄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炉龄预测,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)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炉龄预测的建模,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工业建模具有普遍意义。

参考文献

- [1] 陈念贻等. 计算化学及其应用. 上海:上海科技出版社,1984. 第一章.
- [2] Goldberg D. Genetic algorithm in search, optimization and machine learning, Addison-Wesley Reading, MA, 1989, Chap. 1.
- [3] Davidor Y. Genetic algorithm and robotics, a heuristic strategy for optimization. New Jersey USA: World Scientific Pub. Co 1991.
- [4] Belew K, McInemey J, *et al.* Evolving networks using the genetic algorithm connectionist learning. Artificial life I, Addison-Wesley, 1991. 511-547.
- [5] Koza J. Genetic programming. Cambridge Mass: MIT Press 1992.
- [6] 许立,程兆年等. 化学模式识别要因分析和优化决策方法辅助分析船板钢冲击韧性. 钢铁研究学报,1995,(1): 53-59.
- [7] 许立,程兆年等. 化学模式识别法在钢铁工业中的若干应用. 钢铁研究学报,1993,(4):5-8.

GENETIC PROGRAMMING FOR PROCESS OPTIMIZATION IN INDUSTRY—PREDICTION FOR LIFE OF BASIC OXYGEN FURNACE

CAI YUDONG CHENG ZHAONIAN CHEN NIANYI YANG CHUANREN XU LI

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

LU WENCONG

(Department of Chemistry, Shanghai University, Shanghai 201800)

Abstract In this paper, a genetic programming method for prediction of the life of basic oxygen furnace is presented together with a sample set from a steel plant. The results indicate that the genetic programming seems to be a good method for modeling, and it is possible for this method to become a useful tool for prediction and optimization in iron and steel industry.

Key words process optimization in industry, life of basic oxygen furnace, genetic programming.

蔡煜东 1968年生,1990年毕业于复旦大学,先后在中国科学院上海冶金研究所和上海生物工程中心任助理研究员和副研究员.目前在德国海德堡大学作博士后研究.主要研究兴趣是数据处理方法,仿生学、人工智能和神经网络方法及其工程应用.

程兆年 1942年生,1964年毕业于中国科学技术大学,现任中国科学院上海冶金研究所研究员.主要研究兴趣是物理和工程问题的计算机模拟,数据处理方法,人工智能和神经网络方法及其工程应用.