

运用遗传规划法进行电力系统中长期负荷预测

徐光虎

(上海交通大学电气工程系, 上海 200030)

摘要: 运用遗传规划法进行中长期负荷预测,将预测模型作为遗传规划中的个体,根据“优胜劣汰”的原则,运用复制、变异和交叉三个主要的遗传算子操作,搜索最优预测模型。它根据历史样本数据自动生成负荷预测模型,包括模型的函数形式以及模型参数。同时在模型的实现上对遗传个体进行 Read 线性编码,用十进制编码来代替个体树,通过对编码的操作来实现各种遗传操作,极大地提高了程序运算效率。通过对某地的年用电量进行预测,同时与传统的多元线性回归模型进行比较,结果表明,GP 模型可以显著提高预测精度。

关键词: 电力系统; 负荷预测; 遗传规划; Read 线性编码

中图分类号: TM715 文献标识码: A 文章编号: 1003-4897(2004)12-0021-04

0 引言

1992年,美国学者 Jone R. Koza 将遗传算法应用于计算机程序的优化设计及自动生成,首次提出遗传规划 GP(Genetic Programming)的概念^[1]。GP 的最大特点是采用层次化的结构性语言表达问题,它类似于用计算机程序分行或分段地描述问题。GP 借用生物界进化的规律,根据“优胜劣汰”的原则进行搜索和优化,能够根据环境状态自动改变程序的结构和大小。

用传统的确定性负荷预测方法^[2]进行负荷预测时,事先得确定负荷的变化趋势,然后根据变化趋势来进行曲线拟合,而中长期负荷预测受很多不确定性因素影响,负荷的变化趋势很难事先确定。本文介绍的遗传规划预测方法,可以通过进化算法的具有较高的鲁棒性和全局搜索特性^[3~6],找到负荷变化的最佳回归方程,该方法可避免因主观判断失误而引起的误差,从而提高模型的拟合和预测精度。

1 遗传规划基本原理

1.1 算法表达

遗传规划是用结构层次可变的形式表达问题。在表达式中主要用函数和终止符两类组份。简单地说,终止符表示值,函数表示对值的处理。综合在一起,遗传规划的个体表示对各种值(终止符)的处理过程(函数)。

函数集 F 包含 n 个函数:

$$F = \{f_1, f_2, \dots, f_n\} \quad (1)$$

函数集内的函数 f_i 可以是算术运算符、标准数学函数、布尔运算符等等。

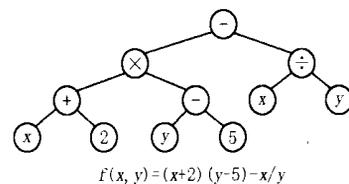
终止符集 T 包含 m 个终止符

$$T = \{t_1, t_2, \dots, t_m\} \quad (2)$$

终止符集内的终止符可以是变量或常量。

为了形象地表达遗传规划的层次结构,通常采用分析树(Parse Tree)的形式,分析树的内节点为函数集 F 中的元素,而其外节点(叶子)为终止符集 T 中的元素,这种分析树常用于表达遗传规划的个体。

任何函数都可用分析树表示,如函数 $f(x, y) = (x+2)(y-5) - x/y$ 可表示为图 1:



$$f(x, y) = (x+2)(y-5) - x/y$$

图 1 $f(x, y)$ 分析树

Fig. 1 $f(x, y)$ analysis tree

1.2 遗传算子

遗传规划如同标准的遗传算法,复制、变异和交叉是三个主要的遗传算子,但是由于其树型结构,它的变异和交叉操作又不同于标准遗传算子的操作。

1) 变异操作

变异操作步骤如下:

在父树上随机地的选择一个节点作为变异点。删除变异点和变异点以下的子树; 随机产生一棵新子树; 将该新子树插入突变点更换旧子树; 如果生成子树的最大深度不大于给定的最大树深度,则使用子树,反之,要么使用父树要么重新进行变异。变异操作如图 2 所示,左侧为父树和一随机产生的子树,右侧为变异后的子树。

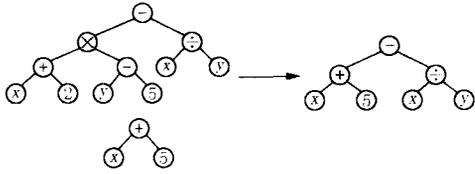


图2 变异操作

Fig. 2 Mutation operation

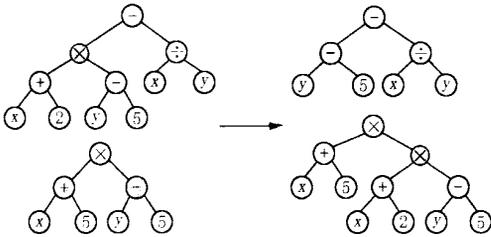


图3 交叉操作

Fig. 3 Crossover operation

2) 交叉操作

交叉操作步骤如下:

在每棵父树上随机地选择一个节点作为交叉点; 交换两个父树的所选交叉点以下的树枝而得到两个子树; 如果所得子树的最大深度不大于给定的最大树深度则选择该子树。交叉操作的过程如图3所示,左侧为父树,右侧为交叉后的子树。

1.3 遗传规划基本步骤

1) 确定个体的表达方式,包括函数集 F 和终止符集 T 。

2) 随机产生初始群体。

初始群体由众多个体组成,每个初始个体都是用随机方法产生。

生成初始个体时,第一步是从函数集 F 中随机选取一个函数作为分析树的根结点。选出根结点后,根据所发出的线数(变量数目),再从函数集 F 和终止符集 T 的并集 $C = F \cup T$ 中按均匀分布的随机方法选出一个元素作为该条线的尾结点。如果选出的仍是函数运算符,则重复执行上述过程;若是终止符,则该分支上的树就停止生长。上述过程从上到下、从左到右不断重复,直至生成一个完整的树为止。在树的生长过程中应注意生成树的深度不要大于给定的最大深度。

3) 计算各个体的适应度。

4) 根据遗传参数进行遗传操作(包括复制、交叉、变异),产生新个体。

5) 反复执行3)及4),直至满足终止条件,取得满意结果。

2 遗传规划负荷预测模型的实现

负荷预测模型的建立,也可看作遗传规划中的符号回归。模型的估计方程相当于遗传规划中的个体,以估计值和实际值之间的偏差作为遗传的驱动力,通过遗传操作,选择最佳拟合方程。遗传规划负荷预测模型的传统实现方法是通过构造分析树来作为个体(即估计方程),所有的遗传操作都是针对树进行的,而树的构造要通过指针和递归算法来实现,这势必会增加算法的复杂度和执行时间,本文采用Read线性编码^[6]来对个体进行编码,通过个体的编码来从函数集或终端集中选择元素构成遗传操作中的个体,这样可提高程序运行效率。以下介绍运用Read线性编码来实现遗传规划负荷预测模型。

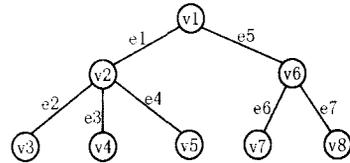


图4 树 $T = T(V, E)$ 的结构图

Fig. 4 Structure of a tree $T = T(V, E)$

2.1 Read 线性编码简介

如图4定义一颗树 $T = T(V, E)$,它由若干顶点(Vertices)和边线(Edges)组成,Read线性编码就是用一组非负整数组成的向量 (a_1, a_2, \dots, a_p) (其中 p 为结点数)按照一定的规则对该树进行编码,以方便编程实现。

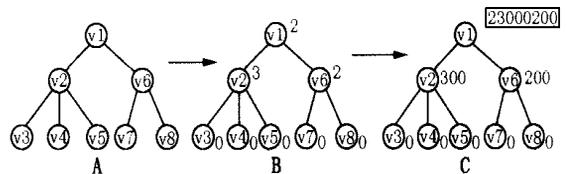


图5 Read 线性编码过程

Fig. 5 Process of Read linear code

如图5显示了图4中树的Read编码过程,A图是初始树;B图对每一个顶点根据其下所接的分支数赋予一个非负整数,该非负整数大小等于分支数大小;C图在树根处显示该树的Read编码向量,它即代表整个树。

Read线性编码树的变异和交叉操作分别如图6、图7所示,在图7中,A、B为两初始个体,对应的A、B为交叉后的个体。

运用GP进行符号回归,各节点符号的选取除从规定的函数集和终止符集中选取外,还可取随机

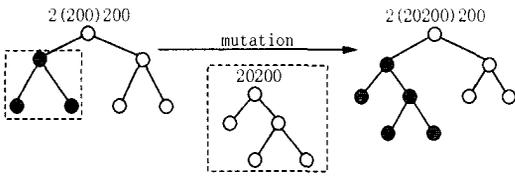


图 6 变异操作

Fig. 6 Mutation operation

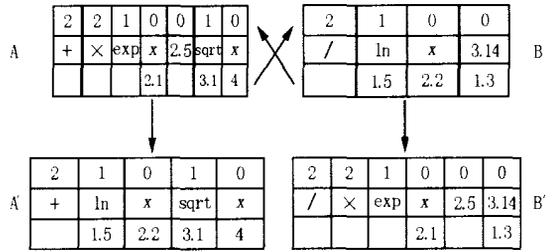


图 9 交叉操作

Fig. 9 Crossover operation

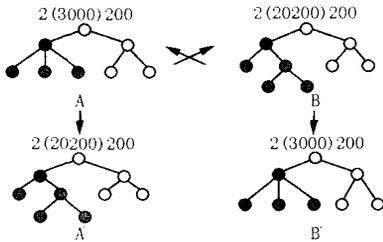


图 7 交叉操作

Fig. 7 Crossover operation

数作为节点的系数,这些系数相当于终止符,这样大大扩充了终止符集的取值范围。表 1 代表一个个体,第一行是个体编码,第二行为每一个编码对应的符号,第三行为节点的系数编码。根据 Read 线性编码规则,表 1 对应的分析树为图 8 所示带系数的树。

表 1 线性编码个体

Tab. 1 Individual for linear code

2	2	1	0	0	2	0	0
+	-	ln	x	v	/	3.0	x
		2.15	-1.23				3.16

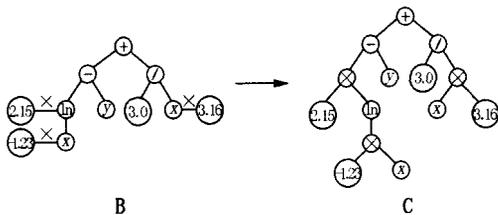


图 8 表 1 所对应的分析树

Fig. 8 Corresponding analysis tree of Tab. 1

图 9 为 A、B 两个个体 Read 编码分别为 2(2100)10 和 2(10)0,两个个体括号部分经交叉操作后形成 A(21010)、B(221000)个体;个体的变异操作较简单,这里就不再列出。

2.2 遗传规划负荷预测流程

程序流程如图 10 所示。其中, Gen 为进化代数, Max 为最大进化代数, M 为群体规模, P_r 、 P_c 、 P_m 分别为复制、交叉和变异概率。

3 算例分析

本文将针对某地 1981~1995 年的二产用电量

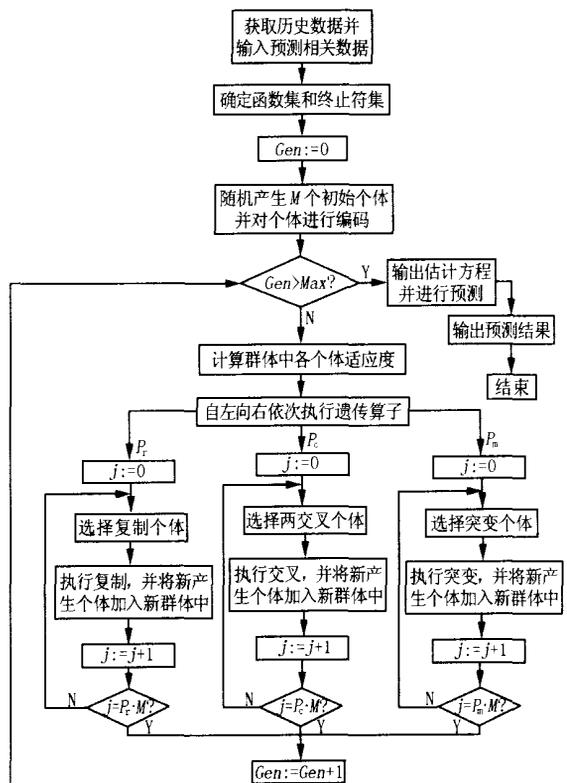


图 10 遗传规划负荷预测流程图

Fig. 10 Flow chart of load forecasting by genetic programming

历史数据,对 1996~2000 年的二产用电量进行预测。为了找到最优的年负荷预测模型,选择合适的函数集和终止符集非常重要,本算例中函数集 $F = \{+, -, \times, \%, \ln, \exp\}$ 。其中 +, -, 和 \times 算子代表普通的加法、减法、乘法;而算子 % 表示“保护性”除法:除以 0 将产生 0 结果而不出错;ln 和 exp 算子代表指数函数和对数函数;终止符集 $T = \{X_1, X_2, X_3\}$,其中 X_1 为二产 GDP、 X_2 为工业用电价格指数比(工业用电价格指数/居民消费价格指数), X_3 为上年二产用电量。

适应度函数取历史电量拟合均方根误差,群体规模为 80,终止代数为 500 代。在同样的运行条件下,采用 Read 线性编码执行时间要比直接采用树的

递归实现编码执行时间短。取适应度最小的一个个体作为最终的 GP 模型,GP 模型与多元线性回归预测结果比较如表 2 所示(电量单位:亿 kWh,限于篇幅仅列出 1991~2000 年数据)。

GP 模型的适应度为 20.4,其估计方程经整理为:

$$Y = X_2^{0.21} \times X_3^{0.84} / \exp(0.95)$$

该方程进行对数变换得: $\ln(y) = 0.95 + 0.21 \ln(X_2) + 0.84 \ln(X_3)$,GP 模型中只保留了 X_2 (工业用电价格指数比)和 X_3 (上年二产用电量)两个变量,使得模型结构更简单。

表 2 GP 模型和多元线性回归模型预测比较

Tab.2 Comparison between GP and regression models

年份	实际值	拟合(预测)值		误差/(%)		
		回归模型	GP 模型	回归模型	GP 模型	
历史 年份	1991	946.82	912.26	919.87	-3.65	-2.85
	1992	1058.71	1013.22	1015.99	-4.3	-4.03
	1993	1154.76	1140.43	1162.41	-1.24	0.66
	1994	1267.4	1250.4	1269.13	-1.34	0.14
	1995	1372.99	1362.89	1348.78	-0.74	-1.76
预测 年份	1996	1449.37	1466.76	1426.99	1.2	-1.54
	1997	1500.52	1554.53	1527.77	3.6	1.82
	1998	1517.68	1616.82	1555.45	6.53	2.49
	1999	1651.3	1660.1	1640.00	0.53	-0.68
	2000	1905.86	1785.76	1885.00	-6.3	-1.09
平均预测误差/(%)				3.63	1.52	

从拟合情况来看,GP 模型的拟合均方根误差为 20.4,而多元线性回归模型拟合均方根误差为 45.8;从预测结果来看,GP 模型的预测误差(绝对值)最大为 2.49%(1998 年)、平均为 1.52%,而多元线性回归模型相应的数值为:6.53%(1998 年)和 3.63%,显然 GP 模型预测精度更高。

4 结语

本文将进化规划算法中的遗传规划方法应用于

电力系统中长期负荷预测,应用结果表明该方法能根据历史负荷与其相关因素的变化关系,自动找出与负荷变化相关的因素并自动生成负荷预测模型,无需事先确定各变量之间的函数关系,从而减小人为判断失误而产生的误差。同时,本文在模型的实现上采用对遗传个体进行 Read 线性编码,运行结果表明这样可提高运行效率。从 GP 模型和多元线性回归模型的比较结果来看,GP 模型结构更简单、预测精度更高。

参考文献:

- [1] Koza J R. Genetic Programming: On the Programming of Natural Selection[M]. MA, MIT Press, 1992.
- [2] 牛东晓,曹树华,赵磊,等(NIU Dong-xiao, CAO Shu-hua, ZHAO Lei, et al). 电力系统负荷预测技术及其应用(The Method and Application of Power System Load Forecasting) [M]. 北京:中国电力出版社(Beijing:China Electric Power Press), 1998.
- [3] 云庆夏(YUN Qing-xia). 进化算法(Evolutionary Algorithm) [M]. 北京:冶金工业出版社(Beijing:Metallurgy Industry Press), 2000.
- [4] Sean L. Two Fast Tree-creation Algorithms for Genetic Programming[J]. IEEE Trans on Evolutionary Computation, 2000, 4(3): 274-283.
- [5] Douglas A A, Helio J C. Symbolic Regression via Genetic Programming[A]. Sixth Brazilian Symposium on Neural Networks, 2000 Proceedings. 173-178.
- [6] Alaa F S, Ahmed M. Forecasting Using Genetic Programming[A]. Southeastern Symposium on System Theory, 2001 Proceedings. 2001, 343-347.

收稿日期: 2003-09-30; 修回日期: 2004-02-27

作者简介:

徐光虎(1974-),男,博士研究生,主要研究方向为负荷预测与建模以及电力系统小信号稳定性研究。

Mid-long term load forecasting in power system by genetic programming

XU Guang-hu

(Department of Electrical Engineering, Shanghai Jiaotong University, Shanghai 200030, China)

Abstract: Genetic programming (GP) is introduced to solve mid-long term load forecasting. Forecasting models are taken as the individuals of GP, which searches the optimal forecasting model by reproduction, mutation and crossover according to the rule of good kept and bad eliminated. It can create automatically load forecasting model including the function form and the numerical coefficients. To realize the model, Read linear code is used to code genetic individually. The tree-like individuals are replaced by decimal codes and the genetic operation is implemented by the operation of the linear code, which improve the computing efficiency. The results of annual forecasting of electric power for some region and comparison with the conventional regression model show that GP model can improve forecasting precision obviously.

Key words: power system; load forecasting; genetic programming; Read linear code