

进化设计中拓扑搜索和参数的协同优化算法

王 斌, 刘德仿

(盐城工学院优集学院, 盐城 224002)

摘要: 为了解决基于遗传编程(GP)的动态系统进化设计过程中拓扑和参数协同优化的问题, 讨论了基于 GP 的进化设计种群拓扑多样性保存策略, 提出了一种拓扑适应度共享-拥挤协同搜索算法。该算法避免计算小生境半径、通过自适应适应度函数来惩罚拓扑子群, 保证了拓扑多样性和阻止局部收敛的发生。实验结果表明, 该算法保证了动态系统进化设计中拓扑和参数同步搜索的平衡, 有效地克服了局部收敛, 能确保获得理想的设计结果。

关键词: 遗传编程; 功率键合图; 进化设计; 拓扑搜索

Cooperative Optimization Algorithm of Topology and Parameter Search in Evolutionary Design Process

WANG Bin, LIU De-fang

(Unigraphics School, Yancheng Institute of Technology, Yancheng 224002)

【Abstract】 To achieve cooperative optimization of a genetic programming (GP)-based dynamic system between structure and parameter during the evolutionary design process, the diversity preservation strategy of evolutionary design topology population is discussed and a topology fitness-sharing & crowding cooperative search algorithm is proposed. This algorithm avoids calculate niche's radius and punish the topology subgroup by self-adaptive fitness-function, thus the diversity of the topology is well preserved and convergence is prohibited. A case is employed to demonstrate the practicality of applying balanced topology and parameter search in evolutionary design of the dynamic system.

【Key words】 genetic programming; bond graphs; evolutionary design; topology search

拓扑开放式进化设计由拓扑进化设计和参数进化设计两部分组成, 因产品结构性能的好坏, 不尽取决于空间拓扑形状, 也取决于零部件的参数及其之间的联结特性, 所以拓扑和参数的进化设计是密不可分的^[1]。然而, 在传统的工程设计领域, 产品拓扑设计是设计者首要考虑的问题, 而参数设计常被置于次要位置。虽然拓扑搜索是进化设计的主要任务, 但是在参数搜索不充分的情况下, 对拓扑性能的评价是不准确的。

实现动态系统进化设计(evolutionary design, ED)的关键是要能够让设计系统精确地、自动地搜索到开放式的拓扑结构(topology)及其相关参数。遗传编程(genetic programming, GP)^[2]以其能够同时进行结构和参数搜索的显著特性成为开放式工程设计领域的重要工具。基于遗传编程进化操作主要分为拓扑搜索和参数搜索两种类型, 其中每一个拓扑搜索问题又包含若干参数搜索子问题。

进化中拓扑搜索和参数搜索之间的主要矛盾如下: (1)过分强调拓扑搜索而忽视参数的充分搜索, 或偏重参数搜索忽视拓扑搜索都是对计算资源的一种严重浪费。(2)如果少数类型的拓扑在整个种群中占相对优势, 会因参数搜索不充分导致新产生的优质结构的适应度值低而无法生存。因此, 本文主要研究如何在拓扑结构搜索和参数搜索间平衡分配搜索能力。

1 原理介绍

目前动态系统的进化设计主要采用GP和功率键合图(bond graphs)相结合的技术^[3,4]。在遗传编程技术中, 程序表示为语法树的形式。遗传编程可以用来生长“树”。这些树代

表一些有组织的指令与函数, 可以用来对给定的初始结构(键合图)进行结构与参数操作, 从而让简单的键合图发育为满足设计目标的设计方案。因为在遗传编程中, 个体的长度是开放式的变长表示, 所以特别适合于键合图的开放式的结构空间搜索^[5]。本文中, GP树实际上代表一个键合图的胚胎程序。键合图是通过胚胎键合图执行一序列的GP函数而生长而成。胚胎键合图通常有几个制定的元件或键被指定为可以修改的“生长点”。GP程序就这些“生长点”为出发点进行发育操作。

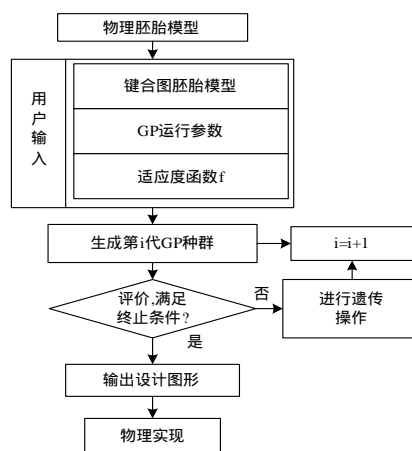


图1 基于GP的进化设计流程

作者简介: 王 斌(1975 -), 男, 讲师, 主研方向: 人工智能, 数据挖掘; 刘德仿, 研究员

收稿日期: 2007-04-20 **E-mail:** ycandy@mail@yahoo.com.cn

整个设计流程如图 1 所示。在开始进化设计前,设计者需要输入以下设计信息:(1)胚胎模型(动力源、载荷等的键合图),或者其他不允许进化修改的系统子模型。(2)遗传编程的运行参数。用于控制初始种群的生成以及进化过程,通常规定初始种群的 GP 树的深度为 3 层~4 层。(3)定义适应度函数(目标函数)。通过执行每一个进化生成的 GP 树,可以对胚胎键合图进行操作形成候选设计方案,对此类候选键合图定义适应度函数用于引导进化向期望的方向发展。在进化过程中的每一阶段,任一候选方案都需经二步评价后赋予一个适应度值。进化过程的终止条件是适应度达到指定值或找到期望的键合图。

2 种群拓扑多样性保存策略

在涉及拓扑变量和参数变量的设计任务中,拓扑和参数间的平衡搜索显得尤其重要。一方面,每一拓扑在经一定次数的参数搜索后,才能较好地开发其性能。这就意味着必须保证种群中的结构个体以一定的量存在。另一方面,又要保证任意少数结构个体都不能在种群中占有绝对优势,否则,结构空间将不能得到充分的开发。

若不能明确区分拓扑搜索和参数搜索之间的不同功能,则易于发生局部收敛。在标准 GP 中,执行交叉(crossover)、变异(mutation)、复制(replication)操作的节点通常是在整个节点集中随机选择的,对拓扑节点和参数的数值节点操作并没有区分,这就意味着新产生的拓扑常会由于其初始参数集的适应度值偏低,而不能繁殖子代。参数集适应度值偏高、结构性能适中的拓扑往往在种群中占有明显的优势,而参数集没有调整到最佳的优质拓扑通常会被遗弃。理论上,只有在参数集得到充分调整的前提下,才能决定该拓扑选用与否。

对于设计方案的性能来说,由于修改拓扑产生的影响远远大于修改某拓扑中的参数产生的影响,因此本文首先通过严格控制拓扑修改节点和参数修改节点的选择比例来平衡拓扑搜索和参数搜索。进化设计中,令拓扑节点的修改选择率为 $P(T \text{ node})=0.2$,参数修改节点选择率 $P(P \text{ node})=0.8$, $T \text{ node}$ 为拓扑修改节点, $P \text{ node}$ 为参数修改节点。

3 算法设计及分析

拓扑适应值共享-拥挤协同搜索算法(fitness-sharing & crowding algorithm, FCA)的出发点是在选择过程和替换过程中均采用相应的策略来维持合理的种群多样性,从而有可能使 GA 既保证收敛又能找到多个峰。该算法的基本内容是在标准遗传算法(genetic algorithms, GA)^[6,7]的基础上,在选择阶段之前,利用适应值共享^[8]的思想进行适应值调整;而在替换阶段,则采用确定性拥挤的思想。与 GA 多样性算法以度量距离来确定峰不同的是,FCA 算法以 GP 中的树结构作为参数和拓扑空间的峰。为维持种群的多样性,FCA 根据下面的适应值调整规则来惩罚个体数量过多的拓扑,使每个拓扑都能进行参数搜索。算法的主要步骤如下。

(1)用适应值共享的方法计算每个拓扑个体的共享后适应值。设 N_{topo} 是个体总数量,代表种群的规模; N_e 是在种群中每一拓扑预期搜索个体的数量; N^{T_i} 是具有相同拓扑 $T_i (i=1,2,\dots,n)$ 的所有个体数量;每一拓扑 T_i 中又包含若干个体 Ind_i 。对每一拓扑 Ind_i 而言,如果 $N^{T_i} > \alpha \cdot N_e$, ($\alpha=0.85$),则 $m^{T_i} = (N^{T_i}/N_e)^\beta$, ($\beta=2$);如果 $m^{T_i} \geq 1$,则 $m > 1$, $f_i' = m \cdot f_i$, f_i 为拓扑个体 T_i 共享前的适应值, f_i' 为调整后的适应值。如果某个物种有较多的个体,那么该物种中所有个体的适应

值将以较大的幅度降低,从而鼓励有较少个体的物种繁衍。

假设拓扑 T_i 中所有个体的适应值均为 f_i , n 代表小生境的数量, r_i 代表拓扑类型 T_i 中个体的数量,则适应值共享的稳定状态可以表示为 $f_i/N^{T_i} = f_j/N^{T_j}$, 其中 $i \neq j$, 且 $\sum_{i=1}^n \sum_{j=1}^n N^{T_i} = N$, 这就意味着在稳定状态下,个体的分布依赖于峰的数量和峰的高度,而不是收敛到唯一的峰上。

(2)采用随机误差比较小的选择方法进行选择过程。

(3)对个体进行交叉和变异操作,产生新个体为 C_Ind_i (其对应的父个体对为 $P_Ind_i \in T_i$)。

(4)在替换阶段采取确定性拥挤的思想来决定下一代中的个体。下面以竞争为例说明具体方法。

1)当 $N_e \geq N^{T_i} > \alpha \cdot N_e$, $N_e > N^{T_j} > \alpha \cdot N_e$ 时,

如果 $P_Ind_i \in T_i$, $C_Ind_i \in T_i$, 且 $f(C_Ind_i) > f(P_Ind_i)$, 则用 C_Ind_i 替换 P_Ind_i , 否则保留 P_Ind_i ;

如果 $P_Ind_i \in T_i$, $C_Ind_i \in T_j$, 且 $f(C_Ind_j) > f(T_j) \cdot m_j$, 则保留 C_Ind_i , P_Ind_i , 否则只保留 P_Ind_i 。

2)当 $N_e \geq N^{T_i} \geq 1$, $\alpha \cdot N_e \geq N^{T_j} \geq 1$ 时,

如果 $P_Ind_i \in T_i$, $C_Ind_i \in T_i$, 且 $f(C_Ind_i) > f(P_Ind_i) > f(T_j) \cdot m_j$, 则用 C_Ind_i 替换 P_Ind_i , 否则保留 P_Ind_i ;

如果 $P_Ind_i \in T_i$, $C_Ind_i \in T_j$, 且 $f(C_Ind_j) > f(T_j)$, 则保留 C_Ind_i , P_Ind_i , 否则只保留 P_Ind_i 。

(3)当 $N_e \geq N^{T_i} \geq 1$, $N^{T_j} = N_e$ 时,

如果 $P_Ind_i \in T_i$, $C_Ind_i \in T_j$, 且 $f(C_Ind_j) > f(T_j) \cdot m_j$, 则用 C_Ind_i 代替 T_j 中 f 值最低的 Ind , 否则只保留 P_Ind_i 。

在计算个体共享后适应值阶段,拓扑适应值共享拥挤算法中无需计算个体之间的距离,并且由于采用顺序或随机的选择方法,使优秀个体的优势在下一代中不能得到完全体现;而适应值共享拥挤算法因其采取随机误差较小的选择方法,可以充分体现优胜劣汰的生物进化原则,因而提高了求解的质量。

4 实验与分析

模拟一个多目标优化系统,选用 5 对目标特征值进行实验(见表 1):

表 1 目标特征值组

特征值数量	特征值 1	特征值 2	特征值 3	特征值 4	特征值 5
5 对	-0.1 ± 5.0j	-1.0 ± 2.0j	-2.0 ± 1.0j	-3.0 ± 0.7j	-4.0 ± 0.4j

实验中采用以下两种不同的进化设计种群。

(1)单一 GP 种群,种群规模:1 000; GP 树最大深度:16;初始深度:4~6;最大节点数:1 000;交叉率:0.9;变异率:0.1;最大繁殖代数:500。

(2)多 GP 子种群,除具有以上单一种群的参数外,还有子群数:10;种群规模:100;迁移间隔:10 代;迁移策略:环状拓扑,即用 10 个最优的个体来代替环中下一子群中 10 个最差的个体。FCA 中: N_{topo} 取 50; $N_e = 1000/N_{topo} = 20$ 。

(下转第 207 页)