

并行化遗传算法研究

陈宝国

(淮南师范学院 数学系, 安徽 淮南 232001)

[摘要] 并行遗传算法(Parallel Genetic Algorithms, PGA)广泛应用于解决各种优化问题。给出了遗传算法并行化目的描述, 做出并行性分析, 详细介绍遗传算法三种结构化并行模型: 踏脚石模型, 岛屿模型, 邻接模型, 最后给出并行遗传算法的硬件支持环境及性能评价。

[关键词] 遗传算法; 并行遗传算法; 并行方法; 模型

[中图分类号] TP301 [文献标识码] A [文章编号] 1009- 9530(2008)03- 0124- 03

1 引言

遗传算法(Genetic Algorithms, 简称 GA)是根据达尔文进化论“优胜劣汰, 适者生存”的一种启发式搜索算法。采用选择、交叉、变异等基本变化算子在解空间同时进行多点搜索, 本身固有并行性。随着大规模并行机的迅速发展, 将并行机的高速性与遗传算法并行性结合起来, 从而促进遗传算法的发展。然而, 仅仅将基本遗传算法硬件并行化会伴随着大量通讯开销等问题, 从而必须对标准 GA 进行改进, 使得并行遗传算法不单单是遗传算法硬件并行实现, 更重要的是结构化的遗传算法。

2 遗传算法并行化目的

遗传算法是模仿自然生物遗传和进化过程中“物竞天择, 适者生存”的原理而开发出的一种多参数、多个体同时优化方法。经过三十多年的开发研究和应用实践, 遗传算法初步显示出了其解决复杂系统优化问题的良好能力, 特别是对一些 NP 组合优化问题的求解, 更表现了优异的性能。如今, 遗传算法已经在旅行商问题的求解、生产调度、图形划分、函数优化、机器学习等众多领域得到了成功的应用, 并显示出良好的性能。标准的遗传算法以个体的集合为运算对象, 对个体所进行的各种遗传操作都有一定的相对独立性, 所以它具有一种天然的并行结构。另一方面, 虽然遗传算法对一个个体编码串的搜索意味着它同时搜索了多个个体模式, 即对个体结构模式的处理具有并行的含义, 但这个并行性是一种隐含的并行性, 其运行过程及实现方法在本质上仍是串行的。这种串行的遗传算法在解决一些实际问题时, 由于它一般具有较大的群体规

模, 需要对较多的个体进行大量的遗传和进化操作, 特别是要对大量的个体进行适应度计算或评价, 从而使得算法的进化运算过程进展缓慢, 难以达到计算速度上的要求, 因而遗传算法的并行计算问题就受到了较大的重视。另一方面, 由于遗传算法的天然并行性, 人们认识到对其进行并行处理的可能性, 从而基于各种并行计算机或局域网, 开发出了多种并行遗传算法。开发并行遗传算法的主要目的是为了提高遗传算法运行速度。开发和应用实践表明, 各种并行遗传算法都能不同程度地达到这个要求。

3 遗传算法的并行性分析

基本遗传算法模型是一个反复迭代的进化计算过程, 通过对一组表示候选解的个体进行评价、选择、交叉、变异等操作, 来产生新一代的个体, 这个迭代过程直到满足某种结束条件为止。对应于基本遗传算法的运行过程, 为实现其并行化要求, 可以从下面四种并行性方面着手对其进行改进和发展。

3.1 并行性 : 个体适应度评价的并行性

个体适应度的评价在遗传算法中占用的运行时间比较大。通过对适应度并行计算方法的开发研究, 可找到并行评价个体适应度的算法, 从而提高个体适应度评价的计算效率。这种并行性的实现可能性及计算效率取决于个体适应度函数的具体表达形式, 依赖于对各种数值并行算法的研究和开发成果。

3.2 并行性 : 整个群体中各个个体适应度评价的并行性

[收稿日期] 2008- 03- 12

[作者简介] 陈宝国(1978-), 男, 安徽安庆人, 淮南师范学院数学系助教, 兰州交通大学硕士研究生, 主要从事数据挖掘及算法分析研究。

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. <http://www.cnki.net>

群体中各个个体的适应度之间无相互依赖关系, 这样各个个体适应度的评价或计算过程就可以相互独立、相互并行地不同的处理机上同时进行, 即不同个体的适应度评价或计算过程可以在不同的处理机上同时进行。

3.3 并行性 : 子代群体产生过程的并行性

在从父代群体产生下一代群体所需进行的遗传运算中, 选择操作只与个体的适应度有关, 这样, 产生子代群体的选择、交叉、变异等遗传操作可以相互独立地并行进行。

3.4 并行性 : 基于群体分组的并行性

从总体上来说, 遗传策法的操作对象是由多个个体所组成的一个群体。从原理上讲, 多个这样的群体应该能共用同一个遗传算法。换一种说法, 同一个遗传算法应该可以同时处理多组群体。这些多组群体可看做是由一个大的群体划分而成的, 若对它们进行进化处理的遗传算法分别置于不同的处理机上, 肯定能够提高运行效率。

在上述四种并行方式中, 前三种方式并未从总体上改变简单遗传算法的特点, 第四种并行方式却对简单遗传算法的结构有较大的改变, 并且这种方式也最自然, 在并行机或局域网环境下实现起来也最简单, 所以受到了人们较大的重视。目前已开发出的并行遗传算法基本上都是基于上述四种并行机制或其组合来实现的。

4 遗传算法的并行化

为提高遗传算法的运算速度、改善其性能, 人们在并行机或局域网环境下开发出了一些并行遗传算法。概括起来, 这些方法大体可分为如下两类: 标准并行方法; 分解型并行方法。

4.1 标准并行方法(standard parallel approach)

这种方法并不改变简单遗传算法的基本结构特点, 即群体中的全部都在统一的环境中进化。其基本出发点是从局部的角度开发个体进化的并行性。在应用遗传算法进行优化计算时, 各个个体的适应度计算、选择、变异等操作是可以相互独立进行的。这样, 利用共享存储结构的并行机, 就可对群体的进化过程进行并行计算以达到提高遗传算法运行速度的目的。这类方法在适应度计算量较大的场合是比较有效的, 上面所介绍的前三种并行性都可以通过这类方法来实现。但另一方面, 由于并行机之间通信等的限制, 选择、交叉、变异等遗传操作的对象集中在一个处理机上较为方便, 所以这类方法的应用受到一些限制, 在有些场合应用效果不太明显。这种并行方法的一个典型例子是由 T.C.Fogarty 等开发的一个基于共享存储结构的并行遗传算法, 该算法将全部群体存放在一个共享的存储器中, 各处理机并行评价各个个体的适应度。

4.2 分解型并行方法 (decomposition parallel ap-

proach)

这种方法是将整个群体划分为几个子群体, 各个子群体分配在各自的处理机或局域网工作站上独立地进行简单遗传算法的进化操作, 在适当的时候各个子群体之间相互交换一些信息。其基本出发点是全局的角度开发群体进化的并行性。这种方法改变了简单遗传算法的基本特点, 各子群体独立地进行进化, 而不是全部群体采用同一机制进化。它是实现上述第 4 种并行性的方法, 并且是一个简单常用、易于实现的方法。这种方法不仅能够提高遗传算法的运算速度, 而且由于保持了各处理机上子群体进化的局部特性, 还能够有效地回避遗传算法的早熟现象。该并行方法有三种不同的群体分组模型: 踏脚石模型、岛屿模型、邻接模型。

(1) 踏脚石模型(stepping-stone model)

这个模型的各个子群体中所含个体的数量多于 1, 各个子群体在其处理机上并行独立地运行简单遗传算法, 子群体之间的信息交换只能是在地理上的邻接处理机之间进行。该模型由于对处理机之间的通信要求不高, 所以实现起来比较简单。图 1 为踏脚石模型的示例。

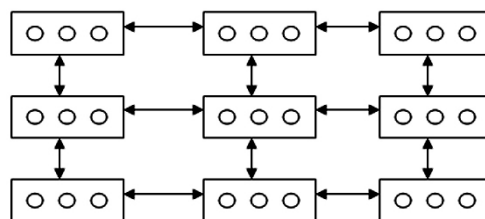


图 1 踏脚石模型

(2) 岛屿模型(island model)

这个模型也叫做粗粒度并行遗传算法 (coarse-grained PGA)。该模型每个处理机上子群体所含个体的数量多于 1, 各个子群体在其各自的处理机上并行独立地运行简单遗传算法, 并且在随机的时间间隔、在随机选择的处理机之间交换个体信息。这种模型适用于 MIMD 机器, 如基于 Transputer 的多处理机系统。图 2 为岛屿模型的示例。

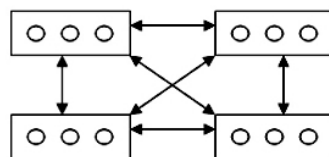


图 2 岛屿模型

(3) 邻接模型(neighborhood model)

这个模型也叫做细粒度并行遗传算法 (fine-grained PGA)。该模型中每个处理机上只分配一个个体, 即子群体只由一个个体组成, 每个子群体只和与其海明距离为 1 的“邻接”子群体相互交换信息。由该模型的特点可知, 即使群体中某一个体的

适合度较高,其作用也仅仅是逐步地才能到其邻近的个体,所以它能够有效地维持群体的多样性,有效地抑制早熟现象。这种模型适用于 SIMD 系统,如阵列式多处理器系统或连接机。图 3 为邻接模型的示例。

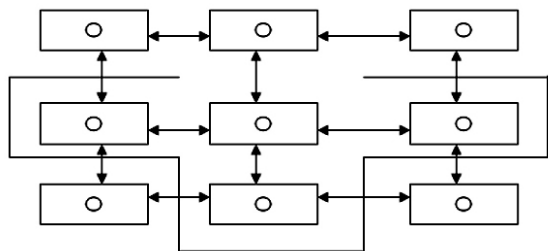


图 3 邻接模型

5 并行遗传算法的硬件支持环境及性能评价

迄今用来实现并行遗传算法的并行机种类很多,既可用粗粒度的并行计算机(如 Transputer 网络),也可用细粒度的并行计算机,既可以在多指令流、多数据流 MIMD 机器上实现,也可以在单指令流、多数据流 SIMD 机器上实现;另外,也可以在局域网环境中实现并行算法。针对某一具体应用问题,到底选用哪种硬件环境,这取决于所使用的并行化方法。为了评价并行算法的性能,人们提出了

很多不同的评价指标,其中最重要的一个评价标准是加速比。设 T_1 为某算法在串行计算机上的运行时间, T_p 是该算法在由 p 个处理机所构成的并行机上的运行时间,则此算法在该并行计算机上的加速比 S_p 定义为: $S_p = T_1 / T_p$; 但对于并行遗传算法而言,由于搜索过程的随机性,仅使用加速比这个指标来衡量其性能的优劣程度是比较困难的。一个比较现实的方法是,设计一些具有不同几何特性的测试函数,通过它们来测量和统计达到最优点时的平均进化代数和平均计算时间,根据这些测试结果来比较不同并行遗传算法的优劣。

参 考 文 献

- [1]陈国良,王熙法,庄镇泉,王东生.遗传算法及其应用[M].北京:人民邮电出版社,1996
- [2]王小平,曹立明.遗传算法—理论、应用与软件实现[M].西安:西安交通大学出版社,2002
- [3]E. G. Talbi and P. Bessiere. A Parallel Genetic Algorithm for the Graph Partitioning Problem[J]. In: Proc. of the 1991 Int. Conf. on Supercomputing. ACM Press, 1991: 312-320

The research of Parallel Genetic Algorithms

CHEN Bao-guo

Abstract: Parallel Genetic Algorithms is used to resolve optimal problems widely. This paper gives the description of using Parallel Genetic Algorithms, puts forward some analysis about parallelism and introduces three structured - parallel model about Parallel Genetic Algorithms: stepping-stone model, island model, neighborhood model. Finally, this article provides supporting environment soft and performance evaluation about Parallel Genetic Algorithms.

Key words: Genetic Algorithms; Parallel Genetic Algorithms; parallel method; model