

并行遗传算法研究综述

高家全, 何桂霞

(浙江工业大学 之江学院, 浙江 杭州 310024)

摘要:扼要概括和评价了并行遗传算法的四种并行模型:主从式模型、粗粒度模型、细粒度模型及混合模型,不仅给出了每种模型的特点及适用范围,而且也指出了每模型在应用中需解决的关键问题.通过模型比较发现粗粒度模型以其较小的通信开销和对种群多样化,获得了最广泛的应用.最后介绍了目前并行遗传算法的主要评价模型,并指出了并行遗传算法的今后主要研究方向,将向理论化、模型化、异步化及混合化等方向发展.

关键词:并行遗传算法;并行计算;并行模型

中图分类号: TP18

文献标识码: A

文章编号: 1006-4303(2007)01-0056-04

A review of parallel genetic algorithms

GAO Jia quan, HE Gui xia

(Zhejiang College, Zhejiang University of Technology, Hangzhou 310024, China)

Abstract: Four kinds of parallel models of parallel genetic algorithms, such as master slave model, coarse grained model, fine grained model and hybrid model, are simply generalized and evaluated. For every model, its characteristics and applicable fields are displayed. Then some key problems which need be solved for improving its application are suggested. Through comparing these models, it is shown that coarse grained model is more widely applied for its little communication overhead and its diversity of population than other models. At the end, some main evaluation models of parallel genetic algorithms are presented. Furthermore, it is shown that researches of parallel genetic algorithms will focus on theorization, modelization, asynchronization etc in the future.

Key words: parallel genetic algorithms; parallel computation; parallel model

0 引言

遗传算法(Genetic Algorithms, GAs)是美国密歇根大学 Holland 教授借鉴生物进化中的“生存竞争”和“优胜劣汰”现象提出来的全局优化算法.从一组初始可行解出发,在不需要除目标函数之外的

其它信息的条件下实现对可行域的全局高效搜索,并以概率 1 收敛到全局最优解^[1].这种良好的特性使它在组合优化领域获得了成功的应用,并成为计算智能领域研究的热点.

随着科学技术的发展,问题的规模不断扩大,复杂度难度增加,对 GAs 求解质量和运行速度都提出了更高的要求,GAs 在处理这些问题时往往都显得

收稿日期: 2006-06-09

基金项目: 浙江省教育厅基金(20051436)

作者简介: 高家全(1972—),男,河南固始人,副教授,博士,研究方向为企业信息化系统与工程、并行数据挖掘、并行生产计划与调度.

“力不从心”。GAs 具有天然的并行性, 非常适合于在大规模并行机上实现, 而随着大规模并行计算机的日益普及, 为并行 GAs 奠定了物质基础。并行遗传算法(Parallel Genetic Algorithms, PGAs)将并行计算机的高速并行性和遗传算法的天然并行性相结合, 并行处理的引入不仅提高了求解速度, 而且由于种群规模的扩大和各子种群的隔离, 使种群的多样性得以丰富和保持, 减少了未成熟收敛的可能性, 提高了求解质量。

1 并行遗传算法的并行模型

实现 PGAs, 不仅要把串行 GAs 等价地变成一种并行方案, 更重要的是要将 GAs 的结构修改成易于并行化的形式, 形成并行模型。目前 GAs 并行模型共分为四类: 主从式模型, 粗粒度模型, 细粒度模型和混合模型。

1.1 主从式模型

主从式模型是遗传算法的直接并行化方案, 不改变遗传算法的基本结构特点, 它只有一个群体, 选择、交叉和变异等全局操作由主节点机(或主处理器)串行进行, 而适应度的评价和计算由各从节点机(或从处理器)并行执行。主节点机(或主处理器)和从节点机(或从处理器)的通信表现在两个方面, 一方面主节点机(或主处理器)把选择的群体部分个体发送给从节点机(或从处理器); 另一方面从节点机(或从处理器)把适应度的评价结果发送给主节点机(或主处理器)。

可以看出, 此模型易于实现, 如果计算时间主要集中在适应度的评估和计算上^[2, 3], 则它将是一个非常有效的并行化方法。另外, 由于它没改变 GAs 的框架结构, 因此可以直接应用 GAs 的理论结果。然而它也有着明显的缺陷, 经常会出现主、从节点机(或处理器)负荷忙闲不均匀, 存在通信瓶颈和通信延迟问题, 并且对于适应度评估量很小情况, 它将失去并行效力。

1.2 粗粒度模型

粗粒度模型又被称作分布式模型或孤岛模型, 是适应性最强和应用最广的遗传算法的并行模型。它将群体依照节点机(或处理器)的个数分成若干个子群体, 各个子群体在各自的节点机(或处理器)上并发独自运行 GAs, 每经过一定的进化代, 各个子群体间将交换若干个个体, 一方面用来引入更优良

的个体, 另一方面丰富了种群的多样性, 达到了防止未成熟早收敛现象^[4]。对于粗粒度模型, 可以看出何时交换个体, 交换哪些个体, 采用什么方式交换是关键问题, 因此目前对粗粒度模型的研究, 主要是解决这几个方面的问题, 也就是我们常说的如何确定迁移规模、迁移策略和迁移拓扑等问题。

1.2.1 迁移拓扑

迁移拓扑确定了子群体之间个体的迁移路径, 目前迁移拓扑大多都采用了类似给定并行机的互联拓扑, 如完全隔离、单向环、双向环、超立方体和网格等。迁移拓扑是影响 PGAs 性能的重要方面, 也是迁移成本的主要因素。在区域数较少时, 可采用全互联的迁移拓扑, 以便迁移个体可以同时迁移到其它各子群体中, 但也有另外一种看法, 在空间上完全隔离对优化处理是有利的。Matsumura T 等在文献[5]中对各种连接拓扑对解的质量影响做了大量对比实验, 发现单向环拓扑虽然收敛速度较慢, 但保证了优良基因在群体间的扩散, 不仅较好地隔离了子群体, 同时也保护了对解的质量影响。为克服收敛速度和解质量之间的矛盾, 提出了一种圆锥形的迁移拓扑, 每间隔一段时间, 圆锥中处于同一层次的小群体间迁移随机挑选的个体, 同时向上层传递最佳个体, 不断保持层内的多样性, 而且也加强了层间的进化压力, 效果更好。

1.2.2 迁移规模

对于粗粒度模型, 确定迁移规模是相当重要的, 因为大规模的迁移将通信时间增多, 而小规模迁移又会形成较小群体的 GAs 独立运行, 影响解的质量。迁移规模由两个重要参数来控制: 迁移周期和迁移率。

迁移周期: 迁移周期决定了各子群体间的迁移间隔, 在通常情况下, 迁移间隔是固定的(此种情况称作同步迁移), 每隔几代迁移一次, 并且迁移率越高, 迁移周期越长。文献[6, 7]通过实验, 发现间隔过大, 会出现早熟收敛现象, 间隔过小, 将使得在可接受的时间内得不到高质量的解。后来人们发现, 如果采用不固定的迁移间隔(此种情况称作异步迁移), 例如当某从节点机(或从处理器)上的子群体出现早熟收敛时或者当子群体的发展缓慢或一直停滞不前时, 才申请迁移。文献[8]提出的异步迁移方式都得到了良好的收敛速度和解的质量。因此迁移时机的确定和采用哪种方式迁移还需作进一步的探究, 迁移周期的确定要根据具体的问题而定。

迁移率:在确定了迁移周期后,紧接着就要确定迁移率.迁移率通常取子群体的百分比或者是给定的指定绝对数.迁移率的确定也需要进一步的研究,因为迁移率过大会破坏子群体的多样性,致使多个搜索进程集中到相同的区域,不利于求解的质量;迁移率过小,将使子群体不能充分利用其它子群体的信息,同样不利于提高解的质量.典型的迁移率是子群体的10%到20%之间.

1.2.3 迁移策略

在确定了迁移周期和迁移率后,进一步就是要确定哪些个体要进行迁移.迁移基本上可以采用与匹配选择和生存选择相同的策略,虽然区域选择更偏向区域内部较好的个体,但在实际的迁移选择和替换中,也可以采用其它标准.

对于迁移个体的选择,大多数情况都采用选择子群体中的最优个体向外移民^[9-10],也可以像匹配选择一种,采用适应度比例或者排列比例来选择迁出个体,也有采用随机选择个体移民的^[11].迁移个体的替换,在大多数情况下都采用使用迁入个体替换子群体中最差个体方法^[9],与迁出的个体选择方法一样,也可以采用适应度比例或者排列比例来选择将被替换的个体,也可以采用随机选择被替换的个体^[11].对于采用何种迁移策略,目前还没有较好的理论依据,大多数情况下都采用最优个体迁出/最差个体被替代方法,但有时随机选择迁出个体或随机被替代个体方法会带来更好的效果.因此迁移策略的研究还需进一步深化.

1.3 细粒度模型

细粒度模型又称作邻域模型,在整个进化过程中虽然保持一个群体,但要求子群体的划分要非常细小,最理想状态是每个节点机(或处理器)只有一个个体,要求各节点机(或处理器)具有极强的通信能力,对于每个染色体,选择和交叉操作都只在所处的节点机(或处理器)及其邻域中进行.由于整个进行过程中,不需要或者需要很少的全局操作,因此充分发挥了GAs并行特性.

对于细粒度模型,采用什么样的邻域结构是需要解决的主要问题之一.邻域结构既决定了子群体个体的空间位置(邻域拓扑),也确定了个体在群体中传播的路径(邻域半径).邻域拓扑确定一个个体的邻居,构成该个体的局部邻域.通常情况下,对于较小的群体中,应采用大范围的拓扑;而对于较大的群体中,多采用具有4个或8个通信邻域的方格超

环面网格.具体采用哪种邻域拓扑较好目前尚无定论,Shapiro B等通过对8个最优近邻域,4个最近邻域在距离 r 内的全部邻域进行实验,发现当 $r > 2$ 时,结果最差,4个邻域模式优于8个邻域模式.

1.4 混合模型

混合模型是近些年快速发展起来的模型结构,主要是通过把前面三种基本模型混合形成层次结构.目前混合模型组合关系主要有三种:粗粒度—细粒度、粗粒度—粗粒度和粗粒度—主从式.在形成的层次结构中,在下层的并行模型中,子群体的规模是真实的,即为一个处理进程所处理的个体数量.而对于上层模型,将每个下层的并行结构都视为一个“集合子群体”,而“集合子群体”之间按上层的并行模型协调运行.对于此混合模型,无论在下层还是在上层,都是子群体内部信息交互量大,之间信息交互量小.目前实际应用中较多的是粗粒度—主从式模型^[12-15].

1.5 模型比较

对于此四种模型,那种模型更好没有一个明确的标准,甚至对于不同问题往往得出互相矛盾的答案.对于主从式模型,由于仅适用计算时间主要集中在适应度评估的问题,因此适用的范围受到了极大的限制;对于细粒度模型,是采用大范围的邻域模型,还是采用小范围直径也有争议,特别是由于对节点机(或处理器)的数量和通信能力要求很高,所以细粒度模型的应用范围也不广;对于粗粒度模型,虽然采用什么样的迁移拓扑、迁移规模和迁移策略还需深入研究,但由于通信开销较小,可获得接近线性的加速比,而且非常适合运行在通信带宽较低的集群系统上^[16],因此得到了广泛的应用;混合模型是在前三种模型的基础上建立起来的,由于具有很好的并行性特点,已成为人们研究的主流,但从应用效果看,目前只有粗粒度—主从式模型应用较好.

2 并行遗传算法的评价模型

对并行遗传算法的评价,大多数都使用Amdahl定律定义的加速比 $s = t(s)/t(p)$.其中 $t(s)$ 为串行遗传算法的执行时间, $t(p)$ 为并行遗传算法的执行时间.由于Amdahl定律适用于负载固定的情况,因此对于遗传算法来说,对主从式和细粒度模型是适用的,在适应度评价计算量较大时,主从式模型可以得到接近线性的加速比^[17].对于细粒度模型由

于使用较少, 因此这方面的研究也较少. 对于评价粗粒度模型, Amdahl 定律仅适用总的群体规模保持不变, 子群体数量和子群体规模成反比的情况, 在这种情况下, 粗粒度模型的加速比可接近线性. 为更好和更实用地评价并行遗传算法, 文献 [18, 19] 结合实际给出了两条评价 PGAs 的指标.

(1) 确定一个适应度指标, 分别利用串行遗传算法和并行遗传算法计算达到某一个个体适用度高于指定适用度指标的时间, 然后让串行搜索时间除以并行搜索时间, 作为 PGAs 的加速比指标.

(2) 给定一定时间, 利用 PGAs 进行优化, 测量能够得到的最高适用度比利用串行 GAs 搜索到的最高适应度高出多少.

对于上面两条评价指标, 不难发现, 第二条评价指标更具有实用价值. 对于不同问题, 由于问题差异、群体大小及采用不同的局部搜索方法等等因素不同, 评价模型得出的结果也会有差异, 因此对于 PGAs 评价模型仍需深一步研究.

3 评注和展望

在前面已扼要地叙述了 PGAs 的现状, 描述了其目前主要的四种计算模型及其评价模型, 并给出了在实际应用中遇到的问题. 从现有遇到的问题来看, PGAs 的研究现在还是相当初步的, 对它进行进一步的研究非常有意义. 从现有研究成果和它的发展趋势来看, PGAs 的研究呈现出如下几个特点:

(1) 关于 PGAs 的理论基础研究. 包括 PGAs 的收敛性, 复杂性分析, 早熟机理探索及预防方法, 群体规模、编码方式及遗传算子等参数设置对算法效率的影响等.

(2) 关于 PGAs 的可执行模型研究. 通过建立并行遗传算法的可执行模型, 来分析遗传算法的优化过程, 研究其遗传参数、迁移拓扑、迁移策略等参数设置方法, 以便在实践中指导并行方式和参数设置.

(3) 关于混合并行遗传算法的研究. 这包括两个方面: 一方面是指朝并行模型混合化方向研究; 另一方面是指在现有模型基础上, 研究如何将遗传算法与局部优化方法结合, 以便有效地提高解的质量.

(4) 关于 PGAs 异步化的研究. 各节点机(处理器)之间以异步方式进行通信, 既节省了同步开销, 又为在 Internet 的分布计算资源上实现并行遗传算法打下了理论基础.

目前, 已有许多学者都致力于 PGAs 的研究, 使得 PGAs 已成功应用到人工神经网络、机器学习、工程计算、复杂组合优化问题等多个领域中, PGAs 有着广阔的应用前景.

参考文献:

- [1] RUDOLPH G. Convergence analysis of canonical genetic algorithm[J]. IEEE Trans, 1994, 5(1): 96 101.
- [2] LOHN J D, COLOMBANS S P. A circuit representation technique for automated circuit design[J], IEEE Trans on EC, 1999, 3(3): 205 219
- [3] GARRETT C A, HUANG J Q, GOLTZ M N, et al. Parallel real valued genetic algorithm for bio remediation optimization of TCE contaminated groundwater[C] // IEEE. Proc CEC99 IEEE. Piscataway, N J: IEEE, 1999: 2183 2189.
- [4] 周远晖, 陆玉昌, 石纯一. 基于克服过早收敛的自适应并行遗传算法[J]. 清华大学学报, 1998, 38(3): 93 95.
- [5] MATSUMURA T, NAKAMURA M, OKECH J, et al. A parallel and distributed genetic algorithm on loosely coupled multiprocessor system [J]. IEICE Trans Fundam Electron Commun Comput Sci, 1998, 81(4): 540 546.
- [6] SUSUMU A, HIDEFUMI S. Effects of migration methods in parallel distributed parameter free genetic algorithm[J]. Electronics and Communication in Japan, 2002, 85(11): 74 80.
- [7] ALBA E, COTTA C, TROYA J M. Numerical and real time analysis of parallel distributed Gas with structured and panmictic populations[C] //IEEE. Proc CEC99 IEEE Piscataway. N J: IEEE, 1999: 1019 1026.
- [8] SALHI A, GLASER H, DE R D. Parallel implementation of a genetic programming based tool for symbolic regression[J]. Information Processing Letters, 1998, 66(6): 299 307.
- [9] GAO J Q. A parallel hybrid genetic algorithm for solving a kind of non identical parallel machine scheduling problems [C] //IEEE Computer Society. Proc 4TH HPC ASIA2006. Beijing: IEEE Computer Society, 2005: 469 472.
- [10] YANG H T, YANG P C, HUANG C L. A Parallel genetic algorithm for deterministic and stochastic labor scheduling problems: implementation on the trans computer networks [J]. IEEE Trans Power Syst, 1997, 12(2): 664 668.
- [11] MATSUMURA T, NAKAMURAM, OKECH J, et al. A parallel and distributed genetic algorithm on loosely coupled multiprocessor system [J]. IEICE Trans Fundam Electron Commun Comput Sci, 1998, 81(4): 540 546.
- [12] 于滨, 姚宝珍, 于艳弘. 一种基于粗粒度一主从式的混合并行遗传算法[J]. 微型电脑应用, 2004, 20(9): 16 18.
- [13] 唐钟, 张葛祥. 一种新并行遗传算法及其应用[J]. 计算机应用与软件, 2005, 22(7): 9 12.
- [14] 高家全. 解决一类家纺企业生产计划排单问题的并行混合遗传算法[J]. 浙江工业大学学报, 2006, 34(4): 402 405.

(下转第 72 页)

开发了相应的仿真实验软件. 非正弦波发生器实验的仿真软件界面如图 6 所示. 图中左边部分是虚拟的非正弦波发生器, 右边部分是仿真的双踪示波器.

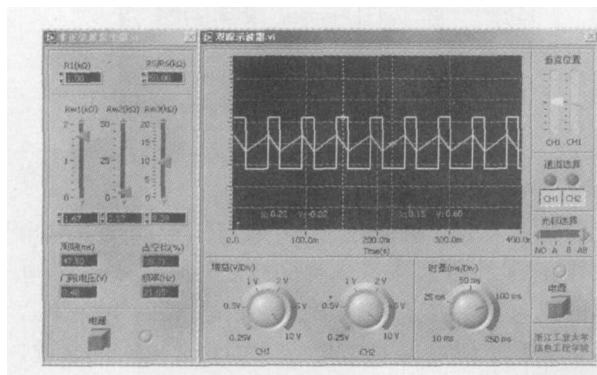


图 6 仿真实验仪器界面

Fig.6 The interface of simulation experiment

4 结束语

电子测试仪器的数字化、计算机化是发展方向,

是电子实验室的最佳选择. 虚拟仪器的发展, 也使我们以不大的代价赶上国外电子实验室的最先进水平成为可能. 通过使用虚拟仪器及仿真仪器软件, 学生可以在相同课时内同时学习电子技术和计算机的使用技巧. 本文开发的数字式双踪存储示波器已在矩形波发生器等电子实验课程中得到应用. 从学生实验报告及反馈意见来看, 数字式双踪存储示波器和仿真仪器软件均获得了良好的效果.

参考文献:

- [1] 程学庆, 房晓溪, 韩薪莘, 等. LabVIEW 图形化编程与实例应用[M]. 北京: 中国铁道出版社, 2005.
- [2] 刘君华. 现代检测技术与测试系统设计[M]. 西安: 西安交通大学出版社, 1999.
- [3] 侯国屏, 王琬, 叶齐鑫. LabVIEW 7.1 编程与虚拟仪器设计[M]. 北京: 清华大学出版社, 2005.

(责任编辑: 陈石平)

(上接第 59 页)

- [15] 高家全, 王雨顺. 解并行多机提前/拖后调度问题的并行遗传算法[J]. 计算机工程与应用, 2006, 42(20): 10-12.
- [16] JOHNSTON W E. Rationale and strategy for a 21st century scientific computing architecture: the case for using commercial symmetric multiprocessors as supercomputers[J]. International Journal of High Speed Computing, 1997, 9(3): 192-222.
- [17] CANTU P E. Designing efficient and accurate parallel genetic algorithms[EB/OL]. [1999-04-07]. <http://www.illigal.ge.uiuc.edu>.
- [18] HAN Y F, ESBENSEN H, SONG L J, et al. Performance improvement of a genetic algorithm for floor planning with parallel computing technology[C]//IEEE. Proc ISCAS97 IEEE. NY: IEEE, 1997: 1544-1547.
- [19] CALEGARI P, GUIDEC F, KUONEN P, et al. Parallel island based genetic algorithm for radio network design[J]. Journal of Parallel Distributed Computing, 1997, 47(1): 86-90.

(责任编辑: 陈石平)