

多重群体遗传算法在装箱问题中的应用研究

李荣

(忻州师范学院 计算机系,山西 忻州 034000)

摘要:装箱问题是一个有很强应用背景的组优化问题,求解极为困难。为有效解决该问题,提出了多重群体遗传算法,给出了具体的遗传算法步骤。在算法中采用新陈代谢的选择策略,以更好地保持进化过程中的遗传多样性。实践表明,引入多重群体遗传算法后,装箱效率有明显的改善和提高。

关键词:多重群体遗传算法;装箱问题;NP-完备;种群

中图分类号:TP301;O221.7

文献标识码:A

文章编号:1673-629X(2007)09-0247-03

Application Study of Multi-Group Genetic Algorithms in Bin-Packing Problem

LI Rong

(Computer Department of Xinzhou Teacher's University, Xinzhou 034000, China)

Abstract: Bin-packing problem is a combinatorial optimization problem having very strong application background and its solution is extremely difficult. In order to solve bin-packing problem efficiently, multi-group genetic algorithm is proposed and the complete algorithm procedures are described. In the algorithm, the selection strategy of metabolism which chooses the individual from multiple generation is used so as to keep the variation of the individual in the process of the evolution. The result of simulation indicates that the efficiency of the packing has improved greatly after using the method of multi-group genetic algorithms.

Key words: multi-group genetic algorithm; bin-packing; NP-completeness; population

0 引言

在计算机科学和工业领域中,装箱问题^[1]有极其广泛的应用背景,包括多处理器任务调度、内存管理、资源分配、运输计划等,特别在现代物流中,很多问题都可以抽象化为装箱问题或其变形,如车辆如何装车,才能充分利用车辆;资金(资源)如何分配,才能得到最大的回报等等,因此对其求解的研究具有广泛的应用价值。

装箱问题虽然广泛存在于许多领域,但从计算复杂性来讲,装箱问题远比一般的NP完全问题更复杂,很难精确求解。到目前为止,人们对装箱问题提出了许多算法,但都有明显不足:穷举法在箱子数目稍大时,存在“组合爆炸”;一些近似算法如:下次适应NF、首次适应FF、降序首次适应FFD、最佳适应BF^[2]等,在解决复杂的装箱问题时,结果与物品的体积数据有

较大关系,在极端情况下很不理想。

遗传算法作为一种基于生物遗传和进化机制的自适应概率优化技术^[3],特别适合求问题的近似最优解。文中针对装箱问题,提出了一种改进的遗传算法,即多重群体遗传算法,并对其进行了求解。

1 装箱问题及数学描述

装箱问题是要将 n 个物品装入许多箱子(最多 n 个箱子)。每个物品有重量 ($w_j > 0$), 每个箱子有重量限制, 每个箱子所装物品的总重量不超过 c_i ($c_i > 0$)。问题是寻找最优的将物品分配到箱子的方案, 从而使每个箱子中物品的重量之和不超过其限制, 而使用的箱子数量最少。该问题可以表示为: 对于 n 个箱子来说, 重量为 w_i 的物品, 其重量限制为 c_i ($i = 1, 2, \dots, n$), w_i 和 c_i 均是正实数。通常, 所有箱子有相同的重量限制 ($c > 0$)。显然, 将每个物品放入一个箱子是一个可行解, 但不是最优解^[4,5]。装箱问题的数学表示如下:

$$\min z(y) = \sum_{i=1}^n y_i \quad (1)$$

收稿日期:2006-11-25

基金项目:山西省忻州师范学院科研基金资助项目(200623)

作者简介:李荣(1974-),女,山西原平人,讲师,硕士,研究方向为中文信息处理、人工智能。

$$s. t. \sum_{j=1}^n w_j x_{ij} \leq c y_i, i \in N = \{1, 2, \dots, n\} \quad (2)$$

$$\sum_{i=1}^n x_{ij} = 1, j \in N \quad (3)$$

$$y_i = 0 \text{ 或 } 1, i \in N \quad (4)$$

$$x_{ij} = 0 \text{ 或 } 1, i, j \in N \quad (5)$$

其中, $y_i = 1$ 表示箱子 i 被装入物品, 反之 $y_i = 0$ 则表示箱子 i 空着。 $x_{ij} = 1$ 表示物品 j 放入箱子 i 中, 反之 $x_{ij} = 0$ 则表示物品 j 未放入箱子 i 中。

装箱问题看似简单, 但从计算机复杂性理论来讲, 装箱问题是 NP-完全问题, 很难精确求解, 目前还没有在多项式时间内求得最优解的算法。

2 多重群体遗传算法模型概述

传统遗传算法模型中每一代只有一个种群, 当求解问题变量数不是太多, 函数关系不是太复杂时, 用这种模型解决装箱问题是一可行思路。但是当随着问题复杂度的增加, 尤其在解群分布不均匀时, 传统方法的进化过程不稳定, 容易出现早熟现象, 陷入局部极优。解决这一问题的方法可考虑对传统遗传算法进行适当改进。

基于以上思路, 文中提出了一种多重群体遗传算法(MGA, Multi-group Genetic Algorithms)^[6]。该算法模仿了自然界生物在进化过程中逐步新陈代谢进而逐步演化的原理, 既结合了传统遗传算法强全局随机搜索能力的优点, 又能提高算法的群体多样性, 保证遗传算法的收敛性。其核心思想是建立一个繁殖群体, 繁殖群体是包含子代群体与父代群体的多层群体, 繁殖群体的规模比种群大得多, 每一代群体并不是只有当代种群, 而是包含了种群与繁殖群体。每一代染色体通过繁殖群体与种群间的新陈代谢来产生下一代个体。多重群体遗传算法模型更好地模拟了大量生物繁殖后代和种群规模相对稳定的进化过程, 也体现了只有少量后代能够存活的竞争选择作用。

3 多重群体遗传算法设计

3.1 适应度函数

算法利用种群中每个个体的适应度值进行搜索, 即对于算法迭代过程中每代个体的优劣程度通过个体适应函数进行评价。因此, 适应度函数的选取至关重要, 它将直接影响到遗传算法的收敛速度以及能否找到最优解。适应度函数值越大, 被遗传到下一代的概率就越大, 解的质量越好。

确定适应度函数的原则是: 最小化使用的箱子数量同时, 尽量装满所有使用的箱子。具体函数如下:

$$f_{\text{eff}} = \frac{\sum_{i=1}^N (F_i/C)^k}{N}$$

式中: N —解中使用的箱子数量; F_i —第 i 个箱子中所有物品的重量之和, 即箱子的填充程度; C —箱子的重量限制; 常数 k ($k > 1$) 表示对装得满的箱子的重视程度, k 越大, 装得满的箱子比一般填充的箱子受到的重视就越大, 这里取 $k = 2$ 。

3.2 编码

编码是应用遗传算法首先要解决的问题, 即把一个问题的可行解从其解空间转换到遗传算法所能处理的搜索空间的转换方法, 它是遗传算法应用成功与否的关键。在遗传算法的实际应用中, 编码就是根据所研究对象的不同性质, 将问题的可行解设计成染色体。由于装箱问题的适应值函数依赖于箱子中物体的群体, 其编码包含两部分: 一部分提供了关于哪个物品属于哪个箱子(群体)的信息; 另一部分对使用的箱子进行编码。例如, 从 1 到 6 对物品进行编码, 染色体物品部分可写作 142352, 染色体的群体部分仅表示箱子, 用字母表示为 ADBCCE。通过查询物品部分, 可知群体是指字母表示的箱子中物品的集合, 即 $B = \{3, 6\}$, $E = \{5\}$, $C = \{4\}$, $D = \{2\}$, $A = \{1\}$, 这种编码使得基因既表示物品又表示群体(箱子), 其物品的群体表示如图 1 所示, 显然, 染色体是变长的。

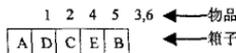


图 1 物品的群体表示

3.3 选择

选择是建立在群体中个体的适应度评估基础之上的。在引进了多重群体之后, 确定选择半径 P_r , 通过种群与繁殖群体之间的新陈代谢对每一代个体进行选择, 即在每一代繁殖的后期, 根据 P_r 的作用范围, 从繁殖群体中选择一定的优秀个体进入种群作为下一代个体。

多重群体遗传算法使用选择算子来对群体中的个体进行优胜劣汰: 适应度较高的个体被遗传到下一代群体中的概率较大; 适应度较低的个体被遗传到下一代群体中的概率较小。文中结合最优保存策略^[7]来保证迄今为止的最优个体不会被交叉、变异等操作所破坏。其具体过程:

(1) 找出当前群体中适应度最高的个体和适应度最低的个体。

(2) 若当前群体中最佳个体的适应度比迄今为止的最好个体的适应度还要高, 则以当前群体中的最佳个体作为新的迄今为止的最好个体。

(3)用迄今为止的最好个体替换当前群体中的最差个体。

3.4 杂交

杂交是指对两个相互配对的染色体以某种方式相互交换部分基因,从而形成两个新的个体,它是产生新的优良基因的主要方法。遗传算法的参数中交叉概率 P_c 的选择是影响遗传算法行为和性能的关键所在,它直接影响算法的收敛性,交叉概率越大,新个体产生速度就越快,然而,遗传算法模式被破坏的可能性也越大,使得具有高度适应度的个体结构很快就会被破坏;如果交叉概率过小,会使搜索过程缓慢,以至停滞不前。本实验中,取交叉概率为 0.7。

由于染色体编码包含 2 部分:箱子和物品的群体,因此,基于这种可变长度的染色体,杂交过程如下:

Step1:选择杂交部分:随机选择两个杂交位置,对每个父代选定杂交部分。

Step2:插入杂交部分:从第 1 个父代插入一些群体(箱子)到第 2 个父代第 1 个杂交位置之前。

Step3:消除空箱子和重复的物品:从产生的后代中原有箱子中去掉空箱子和所有重复出现的物品,使得这些物品原先的从属关系让位于“新”插入的箱子。因消除了一些物品,后代中某些群体不再包含与先前相同的物品。

Step4:重新插入消失的物品:据问题和适应值函数,采用 FFD 局部搜索算法调整新产生的箱子。

Step5:改变两个父代的角色并重新应用第 2 步到第 4 步生成第 2 个子代。

3.5 变异

变异是指将个体染色体编码串中的某些基因座上的基因值用该基因座的其它等位基因来替换,从而形成一个新的基因个体。遗传算法中变异概率 P_m 的选择也是影响遗传算法行为和性能的关键参数,它也直接影响算法的收敛性。如果变异概率太小,就不容易产生新的个体结构;如果变异概率太大,那么遗传算法就变成了纯粹的随机搜索算法。通过反复实验比较,本实验确定变异概率为 0.1。

装箱问题的变异算子针对群体(箱子)进行操作,算子的实现细节依赖于现有的特定群体问题。但可用两条一般性的策略:或启用一个新箱子,或消除一个已使用的箱子。若变异之后解中缺少某些物品,可采用 FFD 启发式方法按随机顺序将其重新放入箱子。

3.6 算法描述

装箱问题的多重群体遗传算法描述如下:

(1)编码形成染色体,并用 FF 启发式方法建立一个初始群体。

(2)计算群体中诸染色体的适应度。

(3)染色体基因的适应与进化。

(a)生成父代。

(b)杂交生成子代。

(c)对子代变异。

(d)评价子代个体。

(e)整理子代个体。

(f)繁殖群体与种群间新陈代谢。

(4)若进化过程已成熟,则统计优化种群、分析进化过程,结束;否则,转(2)。

4 算例研究

文中对一个由 15 个物体和足够多箱子组成的装箱问题进行仿真,采用传统遗传算法,取算法参数 |代数,种群大小,交叉概率,变异概率| 为 |100, 100, 0.7, 0.1| 进行试算,迭代 59 代之后,找到了最优解 4 (最优装箱方案所用箱子数)。而文中引入多重群体遗传算法后,只需 11 代就找到了最优解,并且平均计算时间缩短了大约一半以上,实验结果比较见表 1。

表 1 计算结果比较

求解算法	最小箱子数目	算法收敛到最优点的平均迭代代数	算法的平均计算时间(s)
多重群体遗传算法	4	11	35
传统遗传算法	4	59	72
最优解	4		

5 结束语

装箱问题是一常见而难解的组合优化问题,遗传算法是一种较好的求解方法,但遗传算法在问题规模增大、群体多样性差时,易陷入局部极优。文中针对装箱问题的特点,提出了一种多重群体遗传算法,利用其能够增大遗传群体规模、获得优秀后代的特点,较快地找到了近似最优装箱方案。证明了多重群体遗传算法对装箱问题的求解比传统遗传算法更高效,此改进算法具有很好的全局收敛性,能有效解决装箱问题。

参考文献:

- [1] Kang J, Park S. Algorithms for the variable sized bin packing problem[J]. European Journal of Operational Research, 2003, 147(12):365-372.
- [2] Coffman E G, Garey M R, Johnson D S. Approximation algorithms for bin packing: A survey[C]//Ded H. Approximation Algorithms for NP-Hard Problems. Boston: PWS Publish, 1996:46-93.
- [3] 周 明, 孙树栋. 遗传算法原理及应用[M]. 北京: 国防出版社, 2000.

(下转封三)

③ 压缩快照文件并 FTP 传输方式:

快照代理产生压缩快照文件,再使用 FTP 传输快照文件。

快照建立总时间=快照代理产生压缩快照文件的时间+日志读取器读取事务时间+FTP传输快照文件时间+应用快照到订阅服务器时间。

6.3.2 实验数据

实验数据见表 1。

表 1 不同应用方案下的快照建立时间对比

应用方案	快照建立时间	效率	详细	
默认方式	约为 1 小时 50 分左右	标准		
FTP 传输快照文件方式	约为 1 小时 40 分左右,速度与默认方式相当	稍快		
压缩快照文件并 FTP 传输方式	只需要 1 小时 08 分,速度最快	快	产生压缩快照文件 (Snapshot.cab)	约 33 分 11 秒
			大约是正常情况的两倍多,而这个产生时间与机器的运算能力非常有关,我们发现压缩需要大量消耗 CPU 资源,压缩完的快照文件大约只有不压缩前文件总大小的 1/4 不到	
			FTP 传输压缩快照文件时间	约 17 分钟
			分发代理在订阅服务器上解压缩快照文件时间	约 5 分钟
			应用快照到订阅服务器时间	约 13 分钟

6.3.3 实验结论

目前主要的时间都是消耗在产生压缩快照文件、FTP 传输以及应用快照时间上了。而其中 FTP 传输时间基于网络带宽和数据传输速率,本身可提高的空间有限。而产生压缩快照文件时间和应用快照文件时间则完全取决于机器的运算性能,也就是说,可以通过提高机器设备(只需加大内存和采用高端 CPU,相对于提升网络带宽而言成本更低)的运算能力来进一步降低时间消耗,因而此部署方案是最佳选择,基本可以满足我们期望降低初始化快照的建立时间的目标。同时使用通讯压缩软件可以更进一步降低初始化快照的建立时间。

6.4 如何降低复制数据在网络上传输时间的实验

6.4.1 实验方案

订阅服务器上的分发代理程序 (Distrib.exe) 是通

过 ODBC 连接到分发服务器并读取分发数据库中的事务,然后开始批量传输新的变更数据的。而 ODBC 连接上的数据在网络传输过程中没有采用任何压缩机制,也就是说,对于大量需要传输的数据而言,将会消耗更多的传输时间,因此考虑采用数据压缩软件帮助解决这个问题。方法是在订阅服务器上安装压缩软件客户端,并侦听 1434 端口,让分发代理直接连接到该端口上,这样就相当于将所有分发代理访问分发服务器时在 ODBC 连接上传输的数据内容全部转由通讯压缩软件代为传输,然后在分发服务器端安装通讯压缩软件服务端程序,通讯压缩软件服务端程序配置为连接到分发服务器的 1433 端口(即连接到 SQL SERVER),通过通讯压缩软件的端口转换功能,可将客户端 1434 端口侦听的内容传递到分发服务器的 1433 端口。

6.4.2 实验结论

采用了通讯压缩软件后的传输时间只有不通过通讯压缩软件传输时间的 1/5 左右,亦即传输效率大约提高了 5 倍左右,完全可以满足我们的期望。因此,我们认为图 2 中的方案是目前解决大量复制数据传输的最优部署架构。

7 总结

目前已经有多家公司使用了文中开发的新数据复制管理工具来建立和管理数据复制,并使用通讯压缩技术,采用图 2 中的最优部署架构来建设数据灾备系统。实践证明,文中提出的基于数据复制技术的数据灾备系统是可靠的、高效的和实用的,完全可以满足数据灾备系统的需要。

参考文献:

- [1] 求是科技. Sql Server 2000 数据库管理与开发技术大全 [M]. 北京:人民邮电出版社,2004.
- [2] 袁连海. Sql Server 2000 应用开发实例教程 [M]. 北京:机械工业出版社,2002.
- [3] Nielsen P. Sql Server 2000 Bible [M]. 北京:中国铁道出版社,2004.
- [4] 邹建. 中文版 SQL Server 2000 开发与 管理应用实例 [M]. 北京:人民邮电出版社,2005.
- [5] 龚波. SQL SERVER 2000 教程 [M]. 北京:北京希望电子出版社,2005.

(上接第 249 页)

- [4] 玄光男. 遗传算法与工程优化 [M]. 于歌杰译. 北京:清华大学出版社,2004:48-54.
- [5] 丁香乾. 自适应遗传算法解决集装箱装载问题的方法探讨 [J]. 中国海洋大学学报,2004,34(5):844-848.

- [6] 朱红胜. 多群体遗传算法的特点与应用 [J]. 系统工程理论与实践,1997,6(6):78-85.
- [7] 陈迎春. 约束装箱问题的混合遗传算法求解 [J]. 运筹与管理,2002,11(4):21-25.