

热力学的混合遗传算法及应用

蒋 萍^{1,2}, 唐天兵¹, 李 鼎¹

(1 广西大学 计算机与电子信息学院, 广西 南宁 530004;

2 广西政法管理干部学院 信息工程系, 广西 南宁 530022)

摘 要: 在遗传算法中引入热力学的自由能、熵和内能的概念, 并将传统小生境技术与 Pareto 分级结构结合起来, 设计了一种基于小生境的分层选择策略。结合罚函数和非支配排序的思想, 设计了一种用于约束优化问题的遗传算法。该算法应用到大规模突发事件的应急物资调度分配优化问题中, 所求得的分配优化方案合理有效。

关键词: 热力学; 遗传算法; 自由能; 熵; 物资调度

中图分类号: TN402

文献标识码: A

文章编号: 1000-7180(2013)06-0057-04

Hybrid Genetic Algorithm of Thermodynamics and Application

JIANG Ping^{1,2}, TANG Tian-bing¹, LI Ding¹

(1 School of Computer, Electronics and Information, Guangxi University, Nanning 530004, China;

2 Information Technology Department, Guangxi Administrative cadre Institute of Politics and Law,
Nanning 530023 China)

Abstract: Introducing the concept of thermodynamic free energy and entropy to the genetic algorithm, And traditional niche technology and Pareto hierarchical structure, design stratified selection strategy based on niche, The combined penalty function and non-dominated sorting ideas, design a genetic algorithm for constrained optimization problems. The algorithm is applied to large-scale emergencies, emergency supplies scheduling allocation optimization problems, the obtained distribution optimization is reasonable and effective.

Key words: thermodynamics; genetic algorithm; free energy; entropy; supplies scheduling

1 引言

遗传算法是一种模拟生物进化过程来求解函数优化问题的有效算法, 它不依赖于问题具体的领域, 对各类问题具有很强的鲁棒性, 遗传算法在解复杂问题时也有诸多不足, 如过早收敛, 丢失种群多样性等缺点, 此外, 在解决多目标优化问题时还面临着保持解空间多样性与逼近性平衡的问题^[1]。

文献[2]中利用该热力学定律, 结合热力学熵的概念, 提出一种改进的遗传算法并对多目标函数进行优化计算, 较好地解决了遗传算法过早收敛的问题, 但是所求得的解集在分布性和逼近性方面还存在着不足。文献[3-4]此基础上, 定义等级熵并引入

了分量热力学替换规则, 进行了改进并提高了算法的计算效率^[5]。

在实际应用中, 大规模突发事件中应急物资调度分配问题实际上是一个约束多目标优化问题, 遗传算法解决此类问题通常面临约束处理、早熟、局部搜索能力差、种群多样性的保持, 以及种群多样性和种群逼近度平衡的问题。

本文提出通过热力学的混合遗传算法解决应急物资调度的问题, 结合罚函数和非支配排序的思想设计了约束处理规则, 引入热力学中熵与能量的概念设计热力学算子用来解决局部搜索能力差、种群多样性与逼近性之间平衡的问题, 设计了一种基于分层策略的小生境的策略来保持种群的多样性。通

收稿日期: 2012-07-19; 修回日期: 2012-09-20

基金项目: 国家自然科学基金项目(61063031)

过实验可知,该算法是有效的,提出的方案合理、高效,为这类大规模突发事件中的应急物资调度提供了一种解决方法。

2 算法设计

该优化问题为数值优化问题,采用分配给各厂商的生产配额为编码,生产配额为实数,因此采用这样的编码方式要比采用二进制编码的效率高,且生产配额就是最终要求的方案,因此这种编码方式也避免了计算过程中的编码解码过程。

遗传算法作为一种随机化搜索方法,其种群规模的扩大,将扩大算法搜索范围,对计算起积极作用,但带来计算量的增加,因此不能随意的扩大种群的规模,本文考虑使用初始精英策略,在不改变在计算中的群体规模大小的情况下,扩大搜索范围,其过程是:假设计算中的种群规模为 N ,那么在初始化种群时,随机产生一个规模为 $2N$ 的种群,然后将规模为 $2N$ 的种群通过遗传操作、热力学算子,最后通过分层小生境选择,产生一个规模为 N 的种群,然后将该种群当作初始群体进行后面的遗传迭代计算。

2.1 约束处理

对于约束条件,本文采用罚函数思想,构造罚函数如下:

$$F = \sum_{i=1}^n s_i \quad (1)$$

式中, s_i 为第 i 个约束违反量的绝对值, n 为约束个数。由此可知,罚函数为零的个体为可行解,罚函数大于零的个体为不可行解,罚函数越大说明个体违反约束的程度越大。

通常约束处理采用罚函数法,但是罚函数系数的取值很难把握,在多目标处理时,本文是基于支配排序思想对其进行改进,个体间的罚函数值进行比较,在确定个体之间的支配关系时,罚函数值的比较结果是其参考之一。因此改进后不用给出每个约束惩罚系数,从而避免了因为罚函数系数设定得不合适而影响优化结果。

2.2 多目标处理

进行多目标处理,实际上是在个体具有多个目标函数的情况下,进行个体的优劣判断。如果某一个体的所有目标函数值都分别优于另一个个体的所有目标函数值,那么就定义该个体支配另一个个体。结合罚函数,本文个体之间优劣判断方式如下:

(1)当两个体分别为可行解和不可行解时,那么

可行解优于不可行解;

(2)当两个个体同为可行解时,此时两个个体的罚函数均为零,那么两个个体间按照支配规则进行优劣判断;

(3)当同为不可行解时,罚函数小的个体支配罚函数大的个体,若两个个体罚函数相同,则根据支配规则进行优劣判断。

2.3 热力学设计

交叉算子采用模拟二进制交叉,变异算子采用多项式变异。本文算法中设计了热力学的算子,利用热力学系统中熵与能量的竞争机制来保存种群中适应性与多样性竞争的平衡,遗传算法一个缺点是容易早熟、局部搜索能力较差。对于求解多目标优化问题中,遗传算法在保持种群多样性以及种群多样性与逼近性之间的平衡的问题上也存在着问题,一些学者在遗传算法中引入模拟退火,虽然能在一定程度上解决遗传算法局部搜索能力差的问题。但是,对于遗传算法求解约束多目标问题中,面对的多样性与逼近性的平衡的问题上还是没有提出一个比较好的解决方案。

在热力学系统中,自由能通常定义为: $F = E - HT$,其中该系统自由能为 F ,系统内能为 E ,系统熵值 H ,系统温度为 T ,其中系统的混乱度量是熵。一个与周围环境有热量交换而温度保存不变的热力学系统来说,系统的自由能总是朝着变小的方向进行。因此,系统的内能变小,而熵值变大,那么对于遗传算法的种群,用种群中个体级别的平均值作为系统内能,通过定义个体之间的欧氏距离来计算群体熵值:

$$H = \sum_{k=1}^{i-1} -p_k(i,h) \log p_k(i,h) \quad (2)$$

式中, $P_k(i,h)$ 为群体 P 中的个体 h 和个体 k 之间的欧氏距离与个体数量 n 的比值^[2]。群体熵值大表明群体的多样性好,熵值小表明群体多样性差。显然,熵的增大是受限的,如果太大,那么群体就变成了彻底无序,无法收敛。

最后,群体 P 自由能表示为

$$F = E - \sum_k H_k \quad (3)$$

算子设计概要如下:

(1)参数初始化,相关参数包括:种群 P_0 ,种群规模 N ,温度,扰动系数,扰动次数。

(2)对群体个体执行扰动策略,根据本文选择策略和 metropolis 准则得到种群 P_1 。

(3)新建种群 P_2, P_3 ,其中, P_2 种群 P_0, P_1 进行

合并后的种群, P_3 规模为 N , 初始条件中只有一个个体, 该个体是 P_0 中的最优个体。

(4) 采用式(3)计算 P_2 中的个体分别加入到 P_3 后的自由能, 并进行自由能升序排序。

(5) 如果 P_3 规模达到 N , 则退出, 否则将 P_2 中自由能个体最小的个体加入到 P_3 中, 并将该个体从 P_2 中删掉。

该算子的特点在于: 该算子利用自由能极小化原理并采用扰动策略, 增强了算法的局部搜索能力。以求最小值为例, 种群进化方向是自由能变小, 要求种群自由能变小, 也就是要求最小化种群内能, 最大化种群的熵, 在该算子中, 种群内能即是种群适应度, 种群的熵即是种群的多样性, 那么该算子使得群体进化方向为最小化种群适应度与最大化种群多样性。其中热力学系统中熵与能量的竞争机制将较好的平衡适应度与多样性之间的矛盾。

2.4 分层小生境选择策略

遗传算法求解约束多目标优化问题时, 常常在早期搜索中丢失种群多样性, 而多样性的保持通常采用小生境技术, 而传统 NSGA-II 中的选择操作是通过比较个体的等级与拥挤度来决定个体的优劣, 等级低的个体优于等级高的个体, 同等级的个体, 拥挤度大的个体优于拥挤度小的个体, 这基于目标空间的拥挤度排挤策略在实际计算中有一定概率保留极端解^[5], 因此, 本文基于 NSGA-II 中拥挤度比较算子的思想, 结合小生境技术, 设计了分层小生境选择策略, 通过比较个体的罚函数值以及个体本身的解来对个体进行非支配排序, 依次选择各等级的个体进行小生境选择操作, 并将经过选择的同一个等级个体进行拥挤度计算, 按照拥挤度的由大到小的原则添加到新的种群中, 直到种群满。设种群规模为 N , 其描述如下:

初始化: 将经过快速非支配排序后的种群作为初始种群 P_0 , 种群 $P_1 = \{\}$ 。

For ($i=1; i < \text{种群个体等级的最大值}$ 且 P_1 种群规模 $\leq N; i++$)

对 P_0 中等级为 i 的所有个体进行小生境淘汰操作;

对经过选择操作后的个体进行拥挤度计算并按拥挤度降序排序得到群体 P_2 ;

For ($j=1; P_1$ 种群规模 $\leq N$ 且 $j \leq \text{群体 } P_2 \text{ 的规模}; j++$)

将 P_2 第 j 个个体加入到 P_1 中

End

End

传统小生境淘汰操作是以整个种群为单位进行的, 基于分层策略的小生境操作是根据种群个体的等级进行的, 分别针对每一等级进行淘汰操作, 将留下来的个体根据拥挤度的大小添加到新的种群中去, 当新的种群规模已满时, 则对后面级别的个体不再进行计算, 提高了计算效率, 也改善了单纯依靠拥挤度的排挤策略所带来不足。

2.5 算法流程概述

根据上述介绍, 基于热力学的混合遗传算法流程如下:

(1) 对种群进行初始化。设种群规模为 N , 那么初始化后的种群规模为 $2N$ 。

(2) 对规模为 $2N$ 的种群进行初始精英保留操作, 选择其中 N 个个体作为父代种群 Q 。

(3) 对种群 Q 进行随机联赛操作, 得到种群规模为 $N/2$ 的种群 Q_1 。

(4) 种群 Q_1 进行交叉变异操作后, 得到种群 Q_2 。

(5) 对种群 Q_2 进行自由能极小化计算, 得到种群 Q_3 。

(6) 将 Q, Q_3 合并得到种群 Q_4 。

(7) 对 Q_4 进行改进后的支配排序, 并对其进行分层小生境策略选择操作, 得到规模为 N 的种群 Q_5 。

(8) 判断是否达到最大迭代数, 若没达到, 转到(3), 否则, 输出结果。

3 建立模型

在一般突发事件中, 一般是使用相关储备中心的物资作为应急物资, 当物资供应量紧张时采用一些其他方式如: 市场采购、征用等, 但对于大规模突发事件, 其受灾人口众多, 受灾面积广, 仅仅靠一般的突发事件中的处理手段是不够的, 如在汶川地震后, 帐篷的需求迫切, 需求量也大, 虽然几天内筹集近七万的帐篷到灾区, 但是数量还严重不够, 还需 157 万顶, 对于需求量的缺口, 通过安排相关生产厂家来供应, 于是, 安排哪里厂家生产, 每个生产厂家生产多少, 这都是需要解决的问题, 因为, 每一个生产厂家生产能力、生产成本、生产时间、距离灾区的距离都不同, 因此, 具体安排应该根据这些实际情况而来。

模型基本符号及基本假设:

基本假设:

(1) 受灾地区物资需求量有较确的预测;

- (2) 应急物资的生产厂家是确定的;
- (3) 每个生产厂家运输时间是一个固定值.
- (4) 各生产厂家生产物资与运输物资的成本是一个固定值;

模型中的符号解释如下:

- N : 应急资源的缺量;
- $M_i(i=1,2,3,\cdots,n)$: 应急资源生产商;
- Q_i : 在单位时间内厂商 M_i 的产能;
- X_i : 分配给厂商 M_i 生产的物资数量;
- C_i : 厂商 M_i 的单位物资的生产及运输成本;
- T_i : 从分配给厂商 M_i 任务到物资送达灾区所需的时间.

建模:

$$\min \begin{cases} f_1(x) = \sum C_i x_i \\ f_2(x) = \sum T_i = \frac{x_i}{P_i} + t_i \end{cases}$$
$$\text{s. t.} \begin{cases} \sum x_i = D \\ x_i > 0 \quad (i = 1, 2, \dots, n) \end{cases} \tag{4}$$

式(4)中, 目标函数为: 各厂商从接到分配任务到运送物资到灾区所需要的时间以及各厂商的总成本. 约束条件为: 各厂商生产的物资总量与应急物资需求量相等且各厂商生产量为正数.

4 案例及分析

在某次地震后, 灾区急需帐篷 100 万顶, 但是灾难发生后在 7 天内筹集到的帐篷仅有 30 万顶, 其余的 70 万顶需联合全国的 5 个帐篷生产商生产, 每个厂商生成单位产品的成本、产能均不同, 由于地理原因, 各厂商运送物资到灾区所需要的时间均不同, 其详情见表 1, 为了使帐篷经济的、快速的运到灾区, 应如何分配帐篷生产数量给各厂家呢?

表 1 各厂商单位成本、单位产能、运送物资到灾区的需要的时间

厂商编号	各厂商运送物资到灾区的所耗费的时间	各厂商单位时的生产能力	各厂商生产的单位成本
1	8	3 000	2 500
2	7	2 500	2 850
3	15	2 800	2 630
4	10	1 900	3 200
5	9	3 500	2 180

根据模型和表 2 中的数据可得:

$$\min \begin{cases} f_1(x) = 2\,500x_1 + 2\,850x_2 + 2\,630x_3 + 3\,200x_4 + 2\,180x_5 \\ f_2(x) = \frac{24x_1}{3\,000} + 8 + \frac{24x_2}{2\,500} + 7 + \frac{24x_3}{2\,800} + 15 + \frac{24x_4}{1\,900} + 10 + \frac{24x_5}{3\,500} + 9 \end{cases}$$

$$\text{s. t.} \begin{cases} x_1 + x_2 + x_3 + x_4 + x_5 = 700\,000 \\ x_i \geq 0 \quad i = 1, 2, \dots, 5 \end{cases} \tag{5}$$

表 2 计算结果比较

算法类型	文献[9]	TDGA
x_1	149 611	407 835
x_2	144 224	35 415
x_3	132 902	838
x_3	109 995	333
x_3	163 268	255 579
f_1	$1.842\,5 \times 10^9$	$1.680\,9 \times 10^9$
f_1	$6.278\,6 \times 10^5$	$5.415\,6 \times 10^5$

文献[6]中分配方案为: 厂家 1、2、3、4、5 生产数量分别为: 149 611 顶、144 224 顶、132 902 顶、109 995 顶、163 268 顶. 其总的生产成本和生产及运输的总时间为: $1.842\,5 \times 10^9$ 、 $6.278\,6 \times 10^3$.

本文算法所计算得到的分配方案为: 厂家 1、2、3、4、5 生产数量分别为: 407 835 顶、35 415 顶、838 顶、333 顶、255 579 顶. 其总的生产成本和生产运输总时间为: $1.680\,9 \times 10^9$ 、 $5.415\,6 \times 10^3$.

从表 2 中可知: 厂商 1 和厂商 5 单位成本低且运输耗时少, 因此分配较多的生产任务, 厂商 2 虽然单位成本偏高, 但是由于其运输耗时少, 因此在考虑到应急时间层面上, 对其分配一定的数额, 而厂商 3 和厂商 4 单位成本偏高, 且运输耗时大, 因此, 对其分配的生产任务较少, 由表 2 中数据对比可知, 热力学的混合遗传算法所求得方案更优.

5 结束语

针对大规模突发事件中应急资源调度分配特点, 本文以最小化成本和时间为目标来建立多目标函数模型, 提出了采用热力学的混合遗传算法对多目标模型进行求解, 通过引入热力学中自由能、熵、内能的概念设计了热力学的算子, 该算子既利用扰动策略、热力学自由能极小化原理来加强算法局部搜索的能力, 又利用热力学系统中熵与内能的竞争原理来平衡多目标优化中逼近度和适应度之间的平衡, 在约束处理方面, 将罚函数因子加入到非支配排序中, 对支配算子进行了改进, 避免了因罚函数系数的取值不当而影响计算结果. 在选择策略上, 根据多目标优化过程的特点, 根据种群的 Pareto 分层结构的特点设计了基于分层的小生境选择策略. 最后将算法应用于研究应急物资如何合理的在各厂商之间

(下转第 65 页)

法针对的都是全互联的计算单元,但在实际应用中很多系统都不是这种结构.为了使算法更具有应用价值,我们将对计算单元间任意连接方式的异构系统的任务调度问题进行研究,并在 StarPU 基础上实现一种动态的任务调度平台.

参考文献:

- [1] Kumar R, Tullsen D, Joupp N, et al. Heterogeneous chip multiprocessors[J]. IEEE Computer, 2005(11): 32 - 38.
- [2] Topcuoglu H, Hariri S, Wu M Y, Performance-effective and low-complexity task scheduling for heterogeneous computing[J]. IEEE Trans. Parallel and Distributed Systems, 2002,13(3):260-274.
- [3] Mohammad I Daoud, Nawwaf Kharma. A hybrid heuristic - genetic algorithm for task scheduling in heterogeneous processor networks [J]. Parallel Distrib. Comput., 2011(71):1518 - 1531.
- [4] Mohammad I. Daoud, Nawwaf Kharma, A high performance algorithm for static task scheduling in heterogeneous distributed computing systems [J]. Parallel

Distrib. Comput, 2008(68):399 - 409.

- [5] Yun Wen, Hua Xu, Jiadong Yang. A heuristic-based hybrid genetic-variable neighborhood search algorithm for task scheduling in heterogeneous multiprocessor system[J]. Information Sciences, 2011(8):567 - 581.
- [6] Eswari R, Nickolas S, Path-based heuristic task scheduling algorithm for heterogeneous distributed computing systems[C]// 2010 International Conference on Advances in Recent Technologies in Communication and Computing. India:Kerala,Kottayam,2010.
- [7] Kwok Y K, Ahmad I, Static scheduling algorithms for allocating directed task graphs to multiprocessors[J]. ACM Comput. Surveys, 1999,31(4):406 - 471.
- [8] Sih G C, Lee E A, A compile-time scheduling heuristic for interconnection - constrained heterogeneous processor architectures[J]. IEEE Trans. Parallel Distributed Systems, 1993,4(2):175 - 187.

作者简介:

王川 男,硕士.研究方向为高性能计算.

(上接第 60 页)

进行分配,使得应急物资能更经济、更快速的筹措.实验证明该算法是有效的,所得方案合理、高效.

参考文献:

- [1] Hacer G G, Semra T. A review of applications of genetic algorithms in lot sizing [J]. Journal of intelligent manufacturing, 2010,21(4):575-590.
- [2] 曹先彬,李金龙,王煦法. 热力学遗传算法求解多目标优化问题[C]//1999 年中国智能自动化学术会议论文集(下册). 中国:兰州,1999.
- [3] K F. The free-energy principle: a unified brain theory[J]. Nature Reviews Neuroscience, 2010,11(8): 605.
- [4] 应伟勤,李元香, SHEU Phillip C-Y. 热力学遗传算法

计算效率的改进[J]. 软件学报, 2008,19(7):1613-1622.

- [5] 陈小庆,侯中喜,郭良民,等. 基于 NSGA-II 的改进多目标遗传算法[J]. 计算机应用, 2006,26(10):2453-2456.
- [6] 唐伟勤. 大规模突发事件应急物资调度基本模型研究[D]. 武汉:华中科技大学,2009.

作者简介:

蒋萍 硕士研究生,讲师.研究方向为优化算法.

唐天兵 硕士,副教授.硕士生导师.研究方向为并行分布式计算和计算智能.

李鼎 硕士研究生.研究方向为优化算法.