

# 背包问题的分布估计算法

高尚

(江苏科技大学 计算机科学与工程学院, 江苏 镇江, 212003)

**摘要:** 背包问题经常应用在资源约束条件下的资源分配中, 并在组合数学、计算机科学、复杂性理论和密码学中也有应用。给出了常见的几种背包问题, 将分布估计算法应用于背包问题。分布估计算法提供了一个新的进化模式, 它从有前途的候选解建立概率模型来引导搜索过程。分析个体种群数量、选择比例对算法的影响。仿真结果表明: 分布估计算法求解背包问题是可靠有效的; 此方法具有较好的可扩展性, 修改此算法可解决其他背包问题。

**关键词:** 分布估计算法; 背包问题; 遗传算法

中图分类号: TP301.6

文献标志码: A

文章编号: 1672-7207(2013)S2-0165-04

## Solving knapsack problem by estimation of distribution algorithm

GAO Shang

(School of Computer Science and Engineering, Jiangsu University of Science and Technology, Zhenjiang 212003, China)

**Abstract:** The knapsack problem often arises in resource allocation where there are financial constraints and is studied in fields such as combinatorial mathematics, computer science, complexity theory and cryptography. Several mathematical models of the knapsack problem were proposed and the estimation of distribution algorithms (EDAs) was applied to solve the knapsack problem. Estimation of distribution algorithms (EDAs) offers a novel evolutionary paradigm and makes use of a probabilistic model from the promising solutions to guide the search process. The influence of several parameter design strategies such as population and selection proportion was analyzed. It is concluded that the algorithm with a moderate population and a moderate selection proportion can efficiently find the converged solution among those algorithms. Simulation results show that the EDA is reliable and effective for solving the knapsack problem. This method has good scalability, and it can solve general knapsack problems.

**Key words:** estimation distribution algorithm; knapsack problem; genetic algorithm

背包问题是运筹学中一个典型的组合优化难题, 有着广泛的应用背景, 如货物装载问题、选址问题等。由于此问题比较简单典型, 因此, 评价算法优劣常常以此问题作为的测试对象进行研究。背包问题属于 NP 问题, 目前求解的方法有精确方法(如动态规划、递归法、回溯法、分支限界法等<sup>[1]</sup>), 近似算法(如贪心法<sup>[1]</sup>, Lagrange 法等)以及智能优化算法(如模拟退火算法<sup>[2]</sup>、

遗传算法<sup>[2]</sup>、遗传退火进化算法<sup>[3]</sup>、蚁群算法<sup>[4-5]</sup>)、粒子群优化算法<sup>[6]</sup>。精确方法虽然可以得到准确解, 但时间复杂性与物品数目成指数关系。近似算法和智能优化算法虽然不一定得到准确解, 但可得到比较有效解, 并且时间复杂性比较低。本文作者尝试采用分布估计算法解决此问题。分布估计算法是进化计算领域新兴起的一类随机优化算法, 是当前国际进化计算领

收稿日期: 2013-03-01; 修回日期: 2013-05-02

基金项目: 人工智能四川省重点实验室开放基金(2012RYJ04); “青蓝工程”资助项目(苏教师[2010]27 号)

通信作者: 高尚(1972-), 男, 安徽天长人, 博士, 教授, 从事模式识别与人工智能等方面的研究; 电话: 0511-84408968; E-mail: gaoshang@sohu.com

域的研究热点<sup>[7]</sup>。试验分析表明分布估计算法在求解问题时表现出了比一般遗传算法更好的性能,应用分布估计算法解决工程和科学中的复杂优化问题具有很大的潜力。目前分布估计算法已经在众多领域得到了成功的应用<sup>[8-11]</sup>,如在汽车齿轮机械结构的优化设计、特征选择、不精确图形的模式匹配、软件测试、癌症分类、生物信息学中的特征提取、军事天线的优化设计等方面均有应用,使用分布估计算法解决科学研究和工程应用中的优化问题将是未来几年的研究热点。

## 1 背包问题数学模型

背包问题的描述有许多种形式,最典型的是 0-1 背包问题。0-1 背包问题<sup>[1]</sup>是指:给定  $n$  种物品和一个背包,物品  $i$  的质量是  $c_i$ ,其价值为  $p_i$ ,背包的质量容量为  $C$ ,如何选择物品,使得装入背包中物品的总价值最大。背包问题的解用向量  $X = (x_1, x_2, \dots, x_n)^T$  表示,其数学模型为:

$$\begin{aligned} \max \quad & \sum_{i=1}^n p_i x_i \\ \text{s.t.} \quad & \sum_{i=1}^n c_i x_i \leq C \\ & x_i = 0, 1 \quad (i=1, 2, \dots, n) \end{aligned} \quad (1)$$

带限制的背包问题(bounded knapsack problem)模型是指物品的数量可多取,但受到限制,其模型如下:

$$\begin{aligned} \max \quad & \sum_{i=1}^n p_i x_i \\ \text{s.t.} \quad & \sum_{i=1}^n c_i x_i \leq C \\ & 0 \leq x_i \leq u_i \\ & x_i \text{ 为非负整数} (i=1, 2, \dots, n) \end{aligned} \quad (2)$$

无限制的背包问题(unbounded knapsack problem)模型是指物品的数量可多取,数量不受到限制,其模型如下:

$$\begin{aligned} \max \quad & \sum_{i=1}^n p_i x_i \\ \text{s.t.} \quad & \sum_{i=1}^n c_i x_i \leq C \\ & x_i \text{ 为非负整数} (i=1, 2, \dots, n) \end{aligned} \quad (3)$$

还有许多其他不同类型的背包问题,本文重点讨论模型(1)的解法。

## 2 分布估计算法

分布估计算法的概念<sup>[7]</sup>最初由 Baluja 在 1994 年提出,分布估计算法是一种全新的进化方法。分布估计算法首先构造描述解空间的概率模型,通过对种群的评估,从中选择优秀的个体集合,然后采用统计学习方法根据优秀的个体构造概率模型;然后由这个概率模型随机采样产生新的种群,一般采用随机方法,如此反复迭代,实现种群的进化,直到满足终止条件。分布估计算法根据分布函数的不同而产生了很多算法,这些算法可以分成离散的和连续的两大类,在随机变量之间的依赖关系上又可以分为几类不同的算法。这些算法的基本框架如下:

**Step 1:** 初始化群体,对每一个个体计算适应值;

**Step 2:** 依据适应值,从群体中选择优秀的个体;

**Step 3:** 估计所选择的个体的概率分布;

**Step 4:** 根据分布,采样获得新的个体构成下一代群体,对新个体计算适应值;

**Step 5:** 符合终止条件,算法结束,否则转到 Step2。

算法如图 1 所示。在传统的遗传算法(GA)中,优化问题的候选解用种群个体表示,计算种群中的每个个体的适应值,根据个体适应值进行选择操作,适应值大的个体以较大的概率被选择,然后进行交叉操作和变异等操作,反复进化迭代,直到满足终止条件。而分布估计算法(DEA)没有遗传算法的传统交叉、变异等操作,取而代之的是概率模型的学习。对于遗传

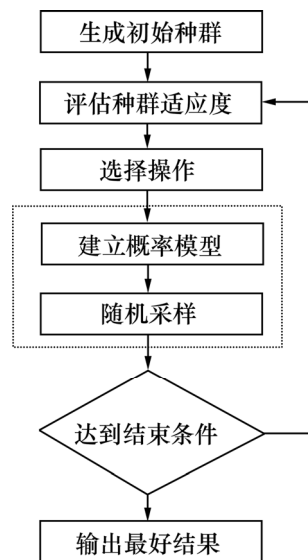


图 1 分布估计算法基本流程

Fig.1 Illustrates flowchart of EDA

算法来说, 其交叉和变异操作有可能会破坏已经进化好的个体, 而分布估计算法利用概率模型和采样两大操作取代了遗传算法中的交叉算子和变异算子, 以一种带有“全局操控性”的操作模式, 可以解决遗传算法存在的这一弊端。而且分布估计算法不需要太多的参数设置, 程序实现比遗传算法简单。

### 3 解 0-1 背包问题的分布估计算法

首先把原约束方程作为罚函数项加入到原目标中, 变成无约束的优化问题, 即

$$\min f = -\sum_{i=1}^n p_i x_i + M \left\{ \min \left\{ 0, \left[ C - \sum_{i=1}^n c_i x_i \right] \right\} \right\}^2 \quad (4)$$

其中:  $M$  为一充分大的正数。

解 0-1 背包问题的分布估计算法如下:

**Step 1:** 以概率  $(p_1, p_2, \dots, p_n)^T = (0.5, 0.5, \dots, 0.5)^T$  随机产生  $N$  个(取 1 的概率为 0.5)个体组成一个初始种群;

**Step 2:** 评估初始种群中所有个体的适应度, 保留最好解;

**Step 3:** 按适应度从高到低的顺序对种群进行排序, 并从中选出最优的  $m$  个个体( $m \leq N$ );

**Step 4:** 分析产生的  $m$  个个体所包含的信息, 估计每个变量取 1 的  $(p_1, p_2, \dots, p_n)^T$ ;

**Step 5:** 从构建的概率模型  $(p_1, p_2, \dots, p_n)^T$  中采样, 得到  $N$  个新样本, 构成新种群;

**Step 6:** 若达到算法的终止条件则结束(如达到规定迭代次数  $n_{\max}$ ), 否则执行 Step 2。

该分布估计算法的时间复杂性估算如下: 以计算适应度操作花费最多, 所以, 时间复杂性大约为  $O(N \cdot n_{\max})$ 。

### 4 数值仿真与分析

采用文献[4]的一个典型的背包问题数据,  $n = 10$ ,  $C = 269$  g,  $\{p_1, p_2, \dots, p_{10}\} = \{55, 10, 47, 5, 4, 50, 8, 61, 85, 87\}$  元,  $\{c_1, c_2, \dots, c_{10}\} = \{95, 4, 60, 32, 23, 72, 80, 62, 65, 46\}$  g。

当  $N = 1\,000$ ,  $m = 0.4N$  时, 统计满足背包的质量容量为  $C$  要求下的最好值与平均值的迭代过程如图 2 所示。迭代一定次数后均收敛到最优解。

影响分布估计算法的性能的主要参数是种群的个

数  $N$  和选出的种群个数  $m$ 。以达到最优目标值为 295 所需要的迭代次数评价标准,  $N$  取不同值,  $m = N/2$ , 各种算法各测试 100 次, 统计数据如表 1 所示。

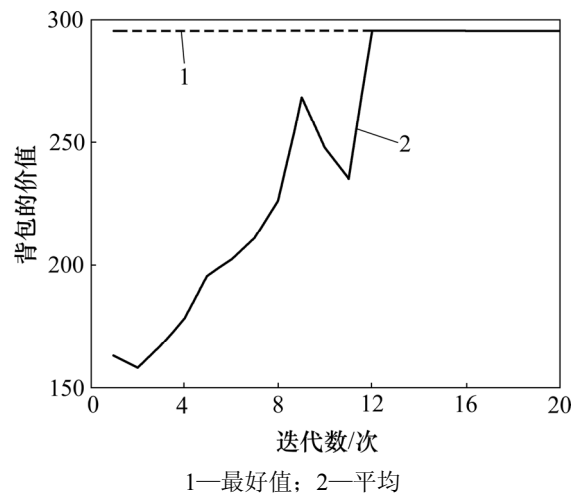


图 2 最好值与平均值的迭代过程

Fig.2 Iterative process of best values and average values

表 1  $N$  取不同值结果比较

Table 1 Comparison results of different  $N$

$N$	平均值迭代数/次	最少迭代数/次	最多迭代数/次
100	—	—	—
200	3.30	1	10
300	2.82	1	6
400	2.53	1	6
500	2.13	1	5
600	1.99	1	5
700	1.71	1	5
800	1.69	1	5
900	1.59	1	4
1 000	1.56	1	4

当  $N = 100$  时, 有时算法进入局部最优解, 达不到全局最优值 295, 无法统计。从表 1 可知:  $N$  越小, 样本数量不足, 效果不好;  $N$  越大, 效果越好, 当然需要的时间也越大。因此,  $N$  取适中即可, 如  $N$  取 800。

当  $N = 800$  时, 选出的种群个数  $m$  占  $N$  的比例不同, 结果也不一样。对于不同的  $m/N$ , 算法各测试 100 次, 统计数据如表 2 所示。

从表 2 可以看出:  $m/N$  越大, 不能体现出选优的优势, 效果越不好; 当然,  $m/N$  太小, 也容易陷入局部极值。  $m/N$  取 10%~30% 时效果比较好。

表 2  $m/N$  取不同值结果比较Table 2 Comparison results of different  $m/N$ 

$(m/N)/\%$	平均值迭代数/次	最少迭代数/次	最多迭代数/次
2.5	—	—	—
5	1.42	1	2
10	1.37	1	2
20	1.53	1	3
30	1.56	1	3
40	1.60	1	4
50	1.69	1	5
60	1.71	1	5

## 5 结论

本文的分布估计算法不仅可以解决背包问题,对于整数规划问题,对该算法作适当修改,也可适用。分布估计算法研究处于初期,还有许多问题值得研究,如算法的收敛性、理论依据等。但从当前的分布估计算法的应用效果来看,这种新型的寻优思想无疑具有十分广阔的前景。

### 参考文献:

- [1] 王晓东. 计算机算法设计与分析[M]. 北京: 电子工业出版社, 2001: 92-168.  
WANG Xiaodong. Design and analysis of computer algorithms [M]. Beijing: Publishing House of Electronics Industry, 2001: 92-168.
- [2] 王凌. 智能优化算法及其应用[M]. 北京: 清华大学出版社, 2001: 17-59.  
WANG Ling. Intelligent optimization algorithms with applications[M]. Beijing: Tsinghua University Press, 2001: 17-59.
- [3] 金慧敏, 马良. 遗传退火进化算法在背包问题中的应用[J]. 上海理工大学学报, 2004, 26(6): 561-564.  
JIN Huimin, MA Liang. Genetic annealing evolutionary algorithm applied to the knapsack problem[J]. Journal of University of Shanghai for Science and Technology, 2004, 26(6): 561-564.
- [4] 马良, 王龙德. 背包问题的蚂蚁优化算法[J]. 计算机应用, 2001, 21(8): 4-5.  
MA Liang, WANG Longde. Ant optimization algorithm for knapsack problem[J]. Computer Applications, 2001, 21(8): 4-5.
- [5] 于永新, 张新荣. 基于蚁群系统的多选择背包问题优化算法[J]. 计算机工程, 2003, 29(20): 75-76, 84.  
YU Yongxin, ZHANG Xinrong. Optimization algorithm for multiple-choice knapsack problem based on ant colony system[J]. Computer Engineering, 2003, 29(20): 75-76, 84.
- [6] 高尚, 杨静宇. 背包问题的混合粒子群优化算法[J]. 中国工程科学, 2006, 8(11): 94-98.  
GAO Shang, YANG Jingyu. Solving knapsack problem by hybrid particle swarm optimization algorithm[J]. Engineering Science, 2006, 8(11): 94-98.
- [7] 周树德, 孙增圻. 分布估计算法综述[J]. 自动化学报, 2007, 33(2): 113-124.  
ZHOU Shude, SUN Zengqi. A survey on estimation of distribution algorithm[J]. Acta Automatica Sinica, 2007, 33(2): 113-124.
- [8] Muhliebe H, Paass G. From recombination of genes to the estimation of distributions. I: Binary parameters[C]// Lecture Notes in Computer Science. Berlin, Germany: Springer Verlag, 1996: 178-187.
- [9] Pelikan M, Godberg D E, Paz E C. Linkage problem, distribution estimation, and Bayesian networks[J]. Evolutionary Computation, 2000, 8(3): 311-340.
- [10] Paul T K, Iba H. Linear and combinatorial optimizations by estimation of distribution algorithms[C]// 9th MPS Symposium on Evolutionary Computation, IPSJ. Kyotanabe, Kyoto, Japan, 2002: 99-106.
- [11] 高尚. 武器-目标分配问题的分布估计算法及参数设计[J]. 东南大学学报: 自然科学版, 2012, 42(S1): 178-181.  
GAO Shang. An estimation of distribution algorithm applied to weapon-target assignment problem and its parameter design[J]. Journal of Southeast University: Natural Science Edition, 2012, 42(S1): 178-181.

(编辑 杨幼平)