

基于二重结构编码遗传算法求解背包问题的研究

刘正龙¹, 杨艳梅², 罗玉军¹

(1. 川北医学院 计算机与数学教研室, 四川 南充 637000;

2. 西华师范大学 数学与信息学院, 四川 南充 637007)

摘要:针对背包问题传统的解决方法有动态规划法、分支界限法、回溯法。传统的方法不能有效地解决背包问题。文中提出二重结构编码的遗传算法解决背包问题,是一种适合于在大量的可行解中搜索最优解的有效算法,在约束条件的处理上结合贪婪算法,既加快了算法的收敛速度,又克服了传统方法容易陷入局部最优的特点,提高了搜索效率。通过计算机仿真试验结果表明,二重结构编码的遗传算法比基本遗传编码有更好的近似解,充分证明了使用二重结构编码的混合遗传算法来求解背包问题的有效性和实用性。

关键词:遗传算法;背包问题;计算机仿真;二重结构编码;二重结构解码

中图分类号:TP301.6

文献标识码:A

文章编号:1673-629X(2013)07-0112-04

doi:10.3969/j.issn.1673-629X.2013.07.028

Study of Solving Knapsack Problem Based on Binary Codes Genetic Algorithm

LIU Zheng-long¹, YANG Yan-mei², LUO Yu-jun¹

(1. Department of Mathematics and Computer, North-Sichuan Medical College, Nanchong 637000, China;

2. Mathematics & Information College, West China Normal University, Nanchong 637007, China)

Abstract:For knapsack problem solutions the traditional methods are the branch and bound method, dynamic programming and backtracking. Traditional methods cannot effectively solve knapsack problem. Propose dual-structure coding genetic algorithm to solve knapsack problem, is an effective algorithm is suitable for a large number of viable solutions for finding optimal solutions, combining with the greedy algorithm on the processing of constraints, not only accelerates the speed of convergence of the algorithm, but also overcomes the traditional methods vulnerable to local optimization features to improve search efficiency. Through the computer simulation results show that dual-structure coding genetic algorithm for approximate solution is better than basic genetic coding, demonstrate using hybrid genetic algorithm for dual-structure codes to solving knapsack problem is effective and practical.

Key words:genetic algorithms; knapsack problem; computer simulation; dual-structure coding; dual-structure decoding

0 引言

旅行包问题(背包问题, Knapsacks Problem), 也是一个典型的 NP (Nondeterministic Polynomial) 完全问题, 背包问题在现实工作中具有广泛的应用, 主要应用于资源分配、投资决策、生物信息、计算分子生物、科学发展观的数学建模等方面。求解背包问题的方法可以归结为启发式算法和最优化算法, 启发式算法分为简单遗传算法求解、贪婪算法求解, 通常用于较大规模的求解, 然而, 最优化算法求解分为迭代算法求解、非线

性动态规划算法求解, 使之能找到解的最优数值解, 但只适用于较小规模的求解^[1]。

文中简单介绍了基本遗传算法和贪婪算法求解旅行包问题的算法设计, 并提出二进制编码的遗传算法研究求解旅行包问题, 最后, 通过计算机实验仿真验证二进制编码的遗传算法设计, 既克服了传统的算法容易陷入局部最优解又加快了该算法的收敛效率, 有更好的近似解^[2]。

收稿日期: 2012-09-08

修回日期: 2012-12-11

网络出版时间: 2013-03-05

基金项目: 四川省教育自然科学基金(12ZB040); 四川省教育发展研究中心基金(CJF10019)

作者简介: 刘正龙(1976-), 男, 副教授, 硕士, 主要研究方向为模式识别与人工智能、计算分子生物学。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20130305.0815.009.html>

1 建立旅行包问题的模型

建立旅行包问题的数学模型实际上是一个动态非线性规划求解问题^[3]。设有 m 个行李要放进旅行包中,用 w_i 表示重量,背包的最大体积容纳重量为 v , p_i ($i = 1, \dots, n$) 表示价值。假设行李 i 被选入旅行包内时,构建函数变量 $x_i = 1$,相反 $x_i = 0$ 。考虑 n 个行李是否放入旅行包内,旅行包内行李的总重量为 $\sum_{i=1}^m w_i x_i$,行李的总价值为 $\sum_{i=1}^m p_i x_i$,怎样解决变量 x_i ($i = 1, 2, \dots, m$) 的值(即确定一个行李的组合)使背包内行李总值为最大。总组合数有 2^m 个,建立数学模型表示:

$$\left. \begin{array}{l} \text{Maximize } \sum_{i=1}^m p_i x_i \\ \text{subject } \sum_{i=1}^m w_i x_i \leq v \\ x_i = 0 \text{ 或 } 1, i = 1, 2, \dots, m \end{array} \right\} \quad (1)$$

在上式中, p_i, w_i, v 均为正数。

通常求解这一问题使用贪婪算法时,将行李价值密度 p_i/w_i 的价值按照降序排,依次将行李放入旅行包内,到超出旅行背包最大容纳体积重量止^[4]。能得到近似最优值解,但是并不能确保一定是最优解。

2 求解背包问题最优解的方法

将求解背包问题的方法可以归结为启发式算法和最优化算法,最优化算法分为递归算法、动态规划法,用于求解较大动态规划问题。从以下几个方面论述基本遗传算法、适应度函数算法、混合遗传算法求解背包问题的设计思想与特点,并提出二重结构编码的遗传算法求解旅行包问题。

2.1 二进制编码简单遗传算法

简单遗传算法是模拟大自然界中生物群在自然环境条件下的进化和繁衍形成的一种自适应全局优化概率寻优算法。简单遗传算法应用于旅行包问题时,使用通常的二进制编码方法,决策变量用一维 $\{x_j (j = 1, \dots, n)\}$ 直接表示 n 位二进制编码字符串,当做个体的遗传基因表示^[5]。在这样的方法中,如果第 j 位基因编码是 1 时,则第 j 个行李被选中;如果第 j 位基因编码是 0 时,则第 j 个行李不被选中。(2) 式是一个 8 变量的旅行包问题:

$$\left. \begin{array}{l} \text{Maximize } 5x_1 + 10x_2 + 13x_3 + 4x_4 \\ \quad 3x_5 + 11x_6 + 13x_7 + 10x_8 \\ \text{Subject to } 2x_1 + 5x_2 + 18x_3 + 3x_4 \\ \quad 2x_5 + 5x_6 + 10x_7 + 4x_8 \leq 25 \\ x_j = 0 \text{ 或 } 1, j = 1, \dots, n \end{array} \right\} \quad (2)$$

产生的 $\langle 1, 0, 0, 1, 0, 1, 0, 1 \rangle$ 是二进制编码的个

体,则该个体对应于选择了行李 1, 4, 6, 8。

处理目标约束条件有两种^[6]:第一种算法是构造目标函数用罚函数法;第二种算法是改进染色体的解码过程结合贪婪算法。第二种方法实际上可以看作是一种混合遗传算法^[7]。

2.2 自适应度函数算法

计算自适应度函数的公式如下:

$$f(x_i) = \begin{cases} \sum_{i=1}^n p_i x_i, & \sum_{i=1}^n w_i x_i \leq v \\ 0, & \sum_{i=1}^n w_i x_i > v \end{cases} \quad (3)$$

本质上,上述(3)式中基于一个考虑违背约束条件的惩罚处理用自适应度函数,如果遇到求解问题规模很大的问题时,即使方法可行,但搜索的效率低下,在很多情况下甚至所得到的结果比贪婪算法的结果还不理想。

2.3 混合遗传算法

染色体解码过程中使用启发式搜索算法“贪婪算法”,具体设计思想简单,对不满足目标约束条件的编码染色体个体,择优编码值为 1 的行李且价值密度较大,到旅行包容量限制且放不下为止,未装入旅行包的行李编码值修正为 0,构造成个体新的染色体编码^[8]。

利用简单的遗传算法求解旅行包问题时,通常只能够求得最优解或近似最优解,适用于求解规模不大的非线性动态规划问题。在先期容易使个别较好个体的后代充当整个生物种群,收敛速度过快的达到某一局部的极小值最优解,很容易造成寻优种群的早熟现象;而在遗传算法的后期,自适应度趋向一致,极优秀的个体在繁衍后代时,优越性很不突出,而使得整个遗传算法种群进化停滞不前。在解决旅行包问题时使用基本遗传算法,揭示了所得解与最优解相差很大甚至低于贪婪算法所求得解,因此需要对基本遗传算法进行改进,使用二进制的编码遗传算法^[9]。

2.4 二重结构编码的遗传算法

将解空间的解数据进行二进制编码,表达为遗传空间的基因型串(即染色体)结构数据,考虑约束条件的满足问题,文中提出了一种改进二进制编码方法(即二重结构编码),提高了遗传算法在区间上收敛的速度。

图 1 所示是二进制编码算法,使用二进制由附加码和变量码两行组成构建种群中个体染色体,上行 $s(i)$ 为函数变量 x_j 的附码 $s(i) = j$,下行为函数变量 $x_{s(i)}$ 对应于附码 $s(i)$ 的数据。

附加码	$s(1)$	$s(2)$...	$s(i)$...	$s(n)$
变量码	$x_{s(1)}$	$x_{s(2)}$...	$x_{s(i)}$...	$x_{s(n)}$

图 1 二进制编码

对个体编码时,首先随机产生附加码 $\{s(i), (i = 1, \cdots, n)\}$,放进第一行;然后随机产生第二行的函数变量码数值(0 或 1),构成个体的二进制编码数据。

种群染色体个体在进行解码时,函数目标约束条件要考虑。如图 2 所示,从左到右的排列方向,函数变量的附码依次要考虑进去,即按排列顺序考虑附码为 $s(i)$ 的行李,然后,进行处理时遇到某个行李违背了函数目标约束条件,强制该行李的函数变量数值 $P_{s(i)}$ 为 0,反之,该行李的变量值 $P_{s(i)}$ 为 1,直到所有行李都处理完为止。

附加码	$s(1)$	$s(2)$...	$s(i)$...	$s(n)$
变量解 码值	$P_{s(1)}$	$P_{s(2)}$...	$P_{s(i)}$...	$P_{s(n)}$

图 2 二进制解码

设计二进制解码算法的流程:

- 第 1 步: $i = 1, \text{sum} = 0$ 。
- 第 2 步:若 $x_{s(i)} = 0$,则 $P_{s(i)} = 0$,到第 4 步执行,否则到第 3 步执行。
- 第 3 步:若 $\text{sum} + a_{s(i)} \leq b, P_{s(i)} = 1$;
 $\text{sum} = \text{sum} + a_{s(i)}, P_{s(i)} = 0$ 。
- 第 4 步: $i = i + 1$,若 $i \leq n$,返回第 3 步,否则停止。

在旅行包问题中的 8 个函数变量,随机产生的附码序列假设为 4,3,8,1,6,2,5,7,则该种群个体的二进制编码为 1,1,0,1,1,0,0,0。它对应于一个可行解,选择了序号 4,3,1,6 的行李。

上述二进制编码,变异和交叉的算子需要重新设计算法,对于交叉用通常的操作算子,种群产生新个体的上行附码会重复出现。假设要很好地解决这个问题采用部分匹配交叉算子^[6]。

图 3 所示为两个子个体 X^*, Y^* 是两个个体 X, Y 经部分匹配交叉算子操作产生的,种群个体的上行附码是少部分对交叉算子操作,种群的子个体下行函数变量码值仍根据其父个体中函数变量码与附码的对应法则确定。

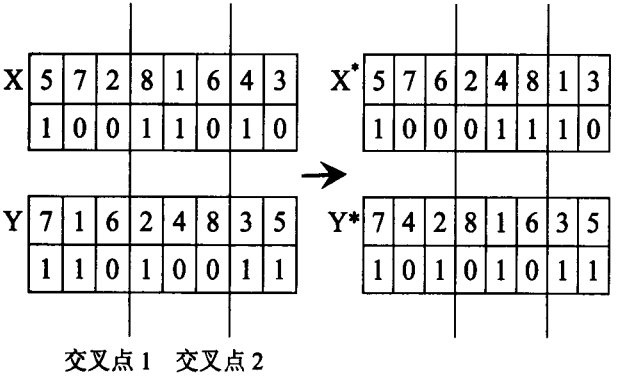


图 3 二进制的 PMX

逆位点遗传算法应用于操作变异。如图 4 所示,两个变异位点是种群父个体随机择优,两位点间的上

行附码按相反顺序重排,下行的函数变量码顺序不作改变^[10]。

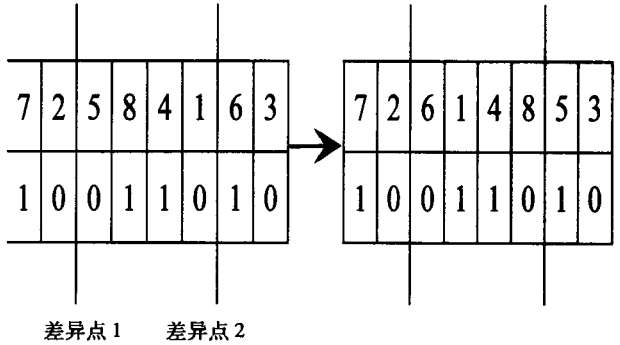


图 4 二进制逆位点变异

3 Matlab 仿真模拟实验

通过计算机仿真模拟将二进制编码遗传算法与基本遗传算法进行比较。求解问题规模为 50 个函数变量,随机地产生 10 个系数值,产生一个旅行包问题的步骤如下:

第 1 步:50 个变量 p_i, w_i 在区间 $[0, 999]$ 内随机地生成。

第 2 步: v 的值按 $A \sum_{i=1}^{50} w_i$ 计算,其中 A 为区间 $[0, 25, 0.75]$ 内一随机数组。

第 3 步:如果所有 $w_i \leq v (i = 1, 2, \cdots, 50)$ 则结束处理;若存在 $w_i > v$,则返回到处理步骤(1),(2)。

表 1 是按照以上步骤产生的 50 个变量的旅行包问题数据。遗传算法处理时种群个体数为 50,操作交叉率为 0.81,变异率为 0.002,停止迭代次数为 600。对每个旅行包问题,将遗传算法通过计算机 MATLAB 模拟仿真计算分块定界法获取的逼近值与末代目标约束函数值比较得相对误差值。

表 1 50 个变量的旅行包问题数据表

	837	12	220	665	359	238	129	887	21	925
	703	599	848	707	117	453	640	620	760	163
p_i	723	383	137	609	260	292	732	882	192	561
	972	879	763	406	985	562	123	61	273	603
	326	447	898	311	421	497	86	822	540	
	828	583	612	34	241	922	794	600	428	590
	256	762	96	985	715	298	423	747	779	384
w_i	163	98	855	446	760	122	895	402	842	872
	734	305	739	628	327	930	9	792	354	92
	769	368	937	488	839	865	486	774	907	508
v	188	337								

对 10 个不同的规划求解问题分别仿真计算,通过

表 2 所示来揭示比较结果。突出地表明同一个规划求解问题,基本遗传算法比二进制编码的遗传算法收敛速度要慢得多。

表 2 SGA 与二进制编码遗传算法计算结果比较

问题(P)	SGA		二重结构编码遗传算法	
	最差值的误差	最优值的误差	最差值的误差	最优值的误差
P ₁	0.076	0.054	0.001	0.000
P ₂	0.128	0.112	0.020	0.000
P ₃	0.130	0.108	0.024	0.009
P ₄	0.140	0.087	0.006	0.044
P ₅	0.037	0.033	0.016	0.000
P ₆	0.145	0.082	0.035	0.000
P ₇	0.195	0.084	0.053	0.037
P ₈	0.082	0.079	0.011	0.000
P ₉	0.173	0.104	0.040	0.020
P ₁₀	0.218	0.133	0.052	0.015

4 结束语

二进制编码下的基本遗传算法在求解背包问题上展示了超强搜索能力,它具有收敛快、搜索速度快的特点,在试验中取得了比基本遗传算法、动态规划、贪心法等更好的求解效果。然而在一般情况下,旅行包问题利用基本遗传算法求解时,得到求解问题结果的近似解也不能满足逼近最优解的要求。二进制编码下的局部搜索寻优算法与基本遗传算法相结合,提高了基本遗传算法寻优方向中的导向,同时也提高了收敛速

度,又克服了传统方法容易陷入局部最优解的特点。通过计算机模拟仿真证明该算法是求解旅行包问题的有效算法,将该算法可以广泛应用于求解其它组合优化问题,比如资源分配、投资决策、科学发展观等优化问题,在并行环境中实现上述算法具有更好的近似解。

参考文献:

[1] Chatterjee S, Carrera C, Lynch A. Genetic algorithms and traveling salesman problems[J]. European Journal of Operational Research,1996,9(3):490-510.

[2] Burkowski F J. Proximity and priority applying a gene expression algorithm to the traveling salesperson problem[J]. Parallel Computing,2004,30(8):803-816.

[3] Katayama K, Sakamoto H, Narihisa H. The efficiency of hybrid mutation genetic algorithm for the traveling salesman problem [J]. Mathematical and Computer Modeling,2000,31(12):197-203.

[4] 张文修,梁怡. 遗传算法的数学基础[M]. 西安:西安交通大学出版社,2001:178-186.

[5] 曾国清. 0-1 背包问题的遗传算法求解[J]. 科技信息,2006(3):242-243.

[6] 霍红卫,许进,保铮. 基于遗传算法的 0/1 背包问题求解[J]. 西安电子科技大学学报,1999,26(4):493-497.

[7] 华福,刘晓路. 面向旅行商问题的一种改进遗传算法[J]. 计算机技术与发展,2011,21(6):51-54.

[8] 刘锐,张金波,刘蕊洁,等. 基于遗传算法求解 0-1 背包问题的算法探讨[J]. 云南民族大学学报(自然科学版),2008,17(4):377-379.

[9] 黄与林. 0-1 背包问题的贪婪算法[J]. 鄂州大学学报,2006,13(6):38-40.

[10] 余建坤,张文彬,陆玉昌. 遗传算法及其应用[J]. 云南民族学院学报(自然科学版),2002,11(4):194-197.

(上接第 111 页)

Springer-Verlag,2004:723-730.

[6] Dong Fangpeng, Akl S G. An adaptive double-layer workflow scheduling approach for grid computing[C]//Proc of the 21st International Symposium on High Performance Computing Systems and Applications. Washington DC:IEEE Computer Society,2007:156-163.

[7] Tretola G, Zimeo E. Activity pre-scheduling in grid workflows [C]//Proc of the 15th Euromicro International Conference on Parallel Distributed and Network-based Processing. Washington DC:IEEE Computer Society,2007:245-253.

[8] 朱敬华,高宏. 无线传感器网络中能源高效的任务分配算法[J]. 软件学报,2007,18(5):1198-1207.

[9] 薛俊芳,邱长华,向东. 在 Pro/E 中自动生成零件拆卸优先约束矩阵[J]. 工程图学学报,2007(3):24-29.

[10] 王歌,张林鎔,贾志新,等. 基于工程语义信息和优先约束矩阵的装配顺序规划方法[J]. 系统仿真学报,2008(19):5262-5267.

[11] de Jong K A. An Analysis of the Behavior of a Class of Genetic Adaptive Systems [D]. Michigan: University of Michigan, 1975.

[12] 梁艳春,冯大鹏,周春光. 遗传算法求解旅行商问题时的基因片段保序[J]. 系统工程理论与实践,2000(4):7-12.

[13] 丁华福,刘晓路,唐远新,等. 面向旅行商问题的一种改进遗传算法[J]. 计算机技术与发展,2011,21(6):51-54.

基于二重结构编码遗传算法求解背包问题的研究

作者：[刘正龙](#)，[杨艳梅](#)，[罗玉军](#)，[LIU Zheng-long](#)，[YANG Yan-mei](#)，[LUO Yu-jun](#)

作者单位：[刘正龙, 罗玉军, LIU Zheng-long, LUO Yu-jun \(川北医学院计算机与数学教研室, 四川南充, 637000\)](#)，[杨艳梅, YANG Yan-mei \(西华师范大学数学与信息学院, 四川南充, 637007\)](#)

刊名：[计算机技术与发展](#) 

英文刊名：[Computer Technology and Development](#)

年，卷(期)：2013, 23(7)

本文链接：http://d.wanfangdata.com.cn/Periodical_wjfz201307028.aspx