

求解大规模 0-1 背包问题的主动进化遗传算法

史亮, 董槐林, 王备战, 龙飞

(厦门大学软件学院, 厦门 361005)

摘要: 针对遗传算法求解大规模 0-1 背包问题中存在的不足, 将定向变异机制引入到遗传算法中, 提出了基于主动进化遗传算法的 0-1 背包问题求解算法。该算法利用概率编码方案对种子个体进行编码, 每代种群中的个体通过对该代种子个体进行测度而产生, 用于定向变异的诱变因子将参与种子个体的进化。实验结果表明, 该算法具有较好的全局寻优能力和执行效率。

关键词: 遗传算法; 定向变异; 0-1 背包问题

Genetic Algorithm Based on Active Evolution for Large Scale 0-1 Knapsack Problem

SHI Liang, DONG Huailin, WANG Beizhan, LONG Fei

(Software School, Xiamen University, Xiamen 361005)

【Abstract】 In order to overcome the problems in resolving large scale 0-1 knapsack problem with genetic algorithm, this paper introduces the directed mutation into the genetic algorithm and presents an active-evolution-based genetic algorithm(AEBGA) for the 0-1 knapsack problem. The algorithm uses probability coding mechanism to construct seed individual, which is used to generate individuals in each generation. In each generation, inducement is generated and used for seed individual evaluation. SGA, GQA and AEBGA are applied to solve large scale 0-1 knapsack problem and experiment results show that AEBGA has good ability of global optimization and high efficiency.

【Key words】 genetic algorithm; directed mutation; 0-1 knapsack problem

0-1 背包问题^[1]是一类非常经典的 NPC 问题。对该问题的研究既有理论价值, 又有实际应用背景, 如资源优化、投资组合等。求解 0-1 背包问题的方法有许多种, 由于该问题的解空间规模同问题输入规模呈指数关系(计算复杂度为 $O(2^n)$), 因此动态规划、回溯法、分支限界法等确定性算法^[1]不适合求解规模较大的 0-1 背包问题。遗传算法^[2]作为一种优化算法, 在大规模 0-1 背包问题的求解上得到广泛的应用, 但遗传算法的全局寻优能力和执行效率往往不理想。文献[3]认为, 遗传算法效率不高的原因是对搜索空间中已经访问的点进行了重抽样。本质上是由其采用的小概率、随机性的不定变异机制所导致的, 这是基于经典进化理论的遗传算法所难以克服的。近年来对遗传算法的改进取得了许多成果, 比如引入量子计算思想, 提出了遗传量子算法^[4], 并将该算法应用于 0-1 背包问题, 取得较为满意的效果, 但由于该算法的变异机制仍沿用传统遗传算法的不定变异机制, 没有从根本上解决遗传算法在求解大规模优化问题时存在的效率和全局寻优能力的问题, 在实际应用中仍不能很好地满足大规模 0-1 背包问题的求解要求。

本文针对遗传算法全局寻优和效率方面的不足, 根据现代生物学引入定向变异, 对遗传算法的变异机制进行了修正, 利用主动进化遗传算法来求解大规模 0-1 背包问题, 以期在保证解的质量的前提下提高整个求解过程的执行效率。

1 主动进化遗传算法

现代生物学的研究成果^[5]表明, 生物体在变异过程中的区域性地位可能并非完全消极被动的, 即不排除生物主动参与其变异发生的过程, 产生有利于其适应环境变化的定向变异。

这种生物体主动参与进化过程、基于定向变异的进化方式被定义为主动进化。那么, 是否可以把这种建立在定向变异基础上的主动进化思想引入到遗传算法领域, 使遗传算法能以一种较为确定、更为有效的方式在解空间中进行寻优, 从而较好地克服标准遗传算法由不定变异导致的盲目性, 提高遗传算法的执行效率。

根据上述分析, 本文对经典遗传算法进行了改进, 提出了一种基于主动进化的遗传算法(active evolution based genetic algorithm, AEBGA)。这种改进的遗传算法取消了交叉操作, 引入了定向变异的变异机制, 具体的实现方式可以根据具体问题进行具体设计^[6,7]。

2 基于主动进化遗传算法的 0-1 背包问题求解方案

0-1 背包问题的实质是从给定的 n 个物品中选择 m 个物品, 在满足背包容量限制的前提下, 使得背包中单位容量的价值量最高。下面将详细介绍用于求解 0-1 背包问题的主动进化遗传算法方案。

2.1 编码

本文采用了文献[7]的编码方案: 利用概率编码取代传统的二值化编码, 即用表示该基因位取 1 的概率值取代原先

基金项目: 国家“985”工程二期基金资助项目(0000-X07204); 福建省自然科学基金资助项目(2006J0222)

作者简介: 史亮(1973-), 男, 博士、讲师, 主研方向: 智能信息处理, 网络信息安全; 董槐林, 教授; 王备战, 博士、副教授; 龙飞, 博士、讲师

收稿日期: 2007-04-28 **E-mail:** sliang@xmu.edu.cn

的 1。将该方案中编码得到的染色体称为种子个体，种子个体在每代产生该代的种群，根据对种群中个体适应度的评估选择该代的最优个体，按定向变异机制完成对种子个体的进化操作。在本文的主动进化遗传算法中，种子个体的数目始终为 1，实际进化的对象就是该种子个体。

2.2 种群的生成

在对种子个体 $R(t) = (r_1^t r_2^t \dots r_n^t)$ ($0 \leq r_i^t \leq 1, i=1,2,\dots,n$ ，其中 n 为物品的数量) 进行测定时，需要根据各个基因位上的具体数值产生用于进行适应度计算的二值化的染色体 $X(t) = (x_1^t x_2^t \dots x_n^t)$ ，其中 x_i^t 为 0 或 1。对于每个 r_i^t ，随机产生一个 (0,1) 之间的随机数 a_i ，如果 $a_i \leq (r_i^t)^2$ ，则 $x_i^t = 1$ ，否则 $x_i^t = 0$ 。每一代根据上述个体生成方案，由该代种子个体产生该代的种群。

2.3 诱变因子的设计

诱变因子的作用是引导个体在进化过程中产生有利于个体适应环境的变异。在求解 0-1 背包问题时，诱变因子为一个固定长度的数字串 $(m_1 m_2 \dots m_n)$ ，其中 $m_i (i=1,2,\dots,n)$ 为 0~1 的小数，表示根据先验知识对物品 i 被选中放入背包的测度。

根据对 0-1 背包问题的分析，该问题的最终目标是在背包容量限制范围内，使背包单位容量的价值量最高。因此在分析物品 i 放入背包的概率 m_i 时，需要考虑 2 个因素：

- (1) 物品 i 的单位重量价值。这里令 $k_i = v_i/w_i - \sum_{i=1}^n v_i/C$ ；
 - (2) 物品 i 的重量同背包容量的比率关系 t_i ，即 w_i/C ；
- 先计算 $q_i = k_i t_i$ ，然后由式(1)得到 m_i ：

$$m_i = (q_i - \min_{1 \leq i \leq n} \{q_i\}) / (\max_{1 \leq i \leq n} \{q_i\} - \min_{1 \leq i \leq n} \{q_i\}) \quad (1)$$

由式(1)可以看出， m_i 是一个 0~1 的小数，表示对第 i 个物品放入背包的测度， m_i 越大，第 i 个物品放入背包的概率越高。诱变因子 M 为 $(m_1 m_2 \dots m_n)$ 。在本遗传算法中， M 为常数因子，在对每代种子个体进行进化操作时， m_i 用于指导种子个体中第 i 个基因位的变异。

2.4 种子个体进化策略

定向变异的操作是通过诱变因子 M 、本代最优个体 $A(t)$ 、历史最优个体 $Y(t)$ 之间的关系来诱发种子个体 $R(t)$ 产生进化的，对 $R(t)$ 上每个 r_i^t 的调整原则如图 1 所示。

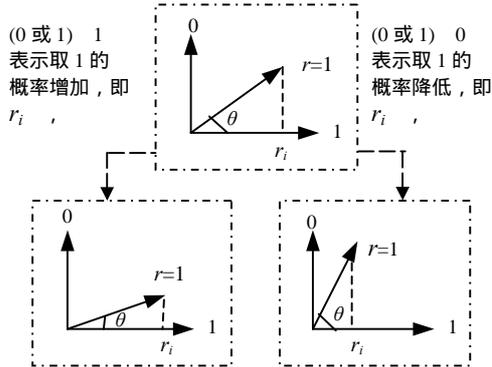


图 1 染色体 $R(t)$ 上每个 r_i^t 的调整原则

这里假定 $A(t)$ 所对应的代价函数值为 $F(A)$ ， $Y(t)$ 所对应的代价函数值为 $F(Y)$ ，图 2 中给出了具体的变异操作， para1 、 para2 和 para3 为调整参数，反映了个体为适应环境进行主动进化的能力，其中 para1 反映基因突变过程，表现在数值上为该基因位上取 1 的概率大幅地上升或下降，因此 para1 的取值较小； para2 和 para3 反映基因渐变过程，表现在数值上

为该基因位上取 1 的概率产生小幅地变化，因此 para2 的取值略大于 para1 ，而 para3 的取值远大于 para1 。

```

i=1;
while(i<=n)
  ki=R(t)中第 i 位上的值 bi
  deltki=acos(ki); //求取 ki 与“1”轴间的夹角 deltki
  //如果本代最优个体 A(t)优于历史最优个体 Y(t)
  if (F(A)>F(Y))
    if (A(t)的第 i 位为 0, Y(t)的第 i 位为 0)
      delta=deltki-deltki/(para_a2*(1+mut(i,1)));
    else if (A(t)的第 i 位为 0, Y(t)的第 i 位为 1)
      delta=deltki-deltki/(para_a1*(1+mut(i,1)));
    else if (A(t)的第 i 位为 1, Y(t)的第 i 位为 1)
      delta=deltki+(pi/2-deltki)/(para_a2*(1-mut(i,1)));
    else if (A(t)的第 i 位为 1, Y(t)的第 i 位为 0)
      delta=deltki+(pi/2-deltki)/(para_a2*(1-mut(i,1)));
    end if
  else //Y(t)优于 A(t)
    if (Y(t)的第 i 位为 0)
      delta=deltki-deltki/(para_a3+mut(i,1));
    else //Y(t)的第 i 位为 1
      delta=deltki+(pi/2-deltki)/(para_a3*(1-mut(k,1)));
    end if
  end if
  i=i+1;
end while

```

图 2 基于诱变因子的定向变异操作

除了 3 个调整参数，本文还将根据先验知识对物品 i 被选中放入背包的测度因素(诱变因子 M)引入到种子个体的进化过程(如图 2)。设计思想如下。

设 $A(t)$ 第 i 个基因位为 a_i ， $Y(t)$ 第 i 个基因位为 y_i ， $R(t)$ 第 i 个基因位为 r_i ，其中 a_i 、 y_i 取值范围都为 0 或 1； r_i 为 0~1 的小数，即第 i 位取 1 的概率。

若 $F(A) > F(Y)$ ，即本代最优个体 $A(t)$ 优于历史最优个体 $Y(t)$ ，则 $R(t)$ 的进化依赖 $A(t)$ 和 $Y(t)$ 在诱导因子的作用下共同完成：

情况 1 如果 $a_i=1, y_i=0$ ， $\text{sita}=\text{sitk}+(\text{pi}/2-\text{sitk})/(\text{para}_a1*(1-\text{mut}(i,1)))$ ，其中 $\text{sitk}=\text{acos}(r_i)$ 。在 0 1 的情况下，如果 $\text{mut}(i,1)$ 越大表明第 i 位取 1 的测度(即 m_i)，则 sita 增加的幅度就越大， $\cos(\text{sita})$ 就越小，该位取 1 的概率就越大，反之则反；

情况 2 如果 $a_i=0, y_i=1$ ， $\text{sita}=\text{sitk}-\text{sitk}/(\text{para}_a1*(1+\text{mut}(i,1)))$ 。在 1 0 的情况下，如果 $\text{mut}(i,1)$ 越大，那么 sita 减少的幅度越小，则 $\cos(\text{sita})$ 就越小，该位取 1 的概率就越大，反之则反。

以上 2 种情况都属于突变类型，控制参数都取 para_a1 。

情况 3 如果 $a_i=1, y_i=1$ ， $\text{sita}=\text{sitk}+(\text{pi}/2-\text{sitk})/(\text{para}_a2*(1-\text{mut}(i,1)))$ ，分析过程与情况 1 类似；

情况 4 如果 $a_i=0, y_i=0$ ， $\text{sita}=\text{sitk}-\text{sitk}/(\text{para}_a2*(1+\text{mut}(i,1)))$ 。分析过程与情况 2 类似。

以上 2 种情况都属于渐变类型，控制参数都取 para_a2 。

若 $F(A) \leq F(Y)$ ，即本代最优个体 $A(t)$ 比历史最优个体 $Y(t)$ 差。此时， $R(t)$ 的进化依赖 $Y(t)$ 和诱导因子 M 完成：

情况 5 如果 $y_i=1$ ，则 $\text{sita}=\text{sitk}+(\text{pi}/2-\text{sitk})/(\text{para}_a3*(1-\text{mut}(i,1)))$ ，分析过程与情况 1 类似；

情况 6 如果 $y_i=0$ ， $\text{sita}=\text{sitk}-\text{sitk}/(\text{para}_a3*(1+\text{mut}(i,1)))$ 。分析过程与情况 2 类似。

以上 2 种情况都属于渐变类型，但进化幅度较情况 3、

情况 4 更为缓慢, 因此控制参数都取 para_a3。

2.5 不定变异机制

需要指出的是, 主动进化遗传算法在引入定向变异机制的同时, 仍保留了不定变异机制。不定变异的具体实现方式同经典遗传算法中相似, 不是直接对种子个体 $R(t)$ 本身进行操作, 而是在产生一个新个体时, 先判断该个体是否有可能发生变异, 如果满足变异概率, 则对新个体随机选择一个基因位并对其进行取反即可。此外, 不定变异概率 P_m 也不再是一个固定值, 而是随着迭代次数的增加而改变, 初始阶段较小, 进入搜索过程后期较大。

2.6 执行步骤

基于诱变因子的主动进化遗传算法的执行步骤如下:

- (1) 初始化种子个体 $R(0)$, 并计算得到诱导因子 M , $t=0$;
- (2) 对 $R(0)$ 进行测定, 得到一组二值化个体 $X(0)$ (即初始种群);
- (3) 对所有个体进行评估, 记录下最佳个体 $A(0)$ 及其适应度值 $F(A(0))$ 。 $Y(0)=A(0)$, $F(Y)=F(A(0))$;
- (4) $t=t+1$;
- (5) 对种子个体 $R(t-1)$ 进行测定, 得到第 t 代的一组二值化个体 $X(t)$;
- (6) 对所有个体进行评估, 记录下最佳个体 $A(t)$ 及其适应度值 $F(A(t))$;
- (7) 依据定向变异机制对 $R(t-1)$ 进行更新, 得到 $R(t)$;
- (8) 如果 $F(Y)<F(A(t))$, $Y(t)=A(t)$, $F(Y)=F(A(t))$;
- (9) 如果满足结束条件, 则终止, 否则, 调整 P_m 并继续执行步骤(4)~步骤(9)。

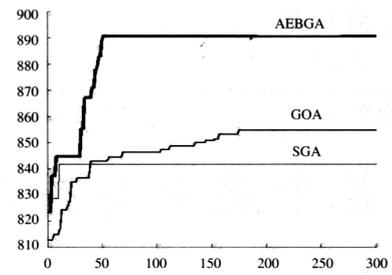
3 实验结果与分析

实验中数据的产生参照文献[4]。设 $w_i \in [1,10]$ 的随机数, $v_i = w_i + l_i$, $l_i \in [0,0.5]$ 的随机数, 包容量 $C = \frac{1}{2} \sum_{i=1}^n w_i$, 生成物品数 n 分别为 200、500 和 1 000 的 3 组数据。为了进行对比分析, 本文对相同的数据样本采用标准遗传算法 SGA、遗传量子算法 GQA 和本文的主动进化遗传算法 AEBGA 各测试了 20 次, 3 种算法中种群规模都为 80, 进化代数都为 300。实验中, SGA 中的交叉概率 0.85, 变异概率 0.007 5。AEBGA 中的 3 个控制参数 para1、para2 和 para3 的取值分别为 2.5、8 和 40。GQA 的设定参考文献[4]。实验在 CPU 为 P4 2.8GHz、内存为 1GB 的微机, 利用 Matlab7.0 进行编程实现的。

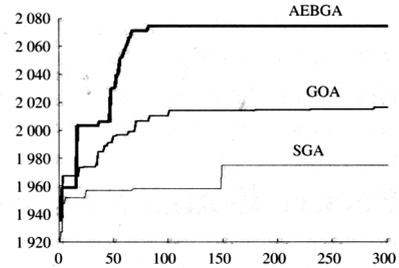
表 1、图 3 分别给出了物品数为 200、500 和 1 000 的实验结果。通过比较可以发现, 本文的主动遗传算法无论是解的质量还是求解的效率都优于遗传量子算法, 较之标准遗传算法有显著提高。

表 1 3 种算法的实验结果

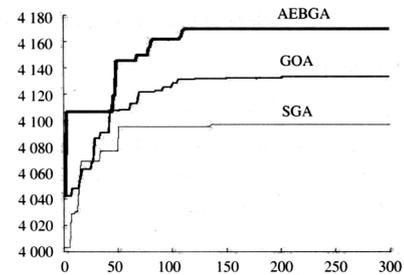
物品数目	算法	20 次实验中最优个体的适应度			执行时间/s
		最好	最差	平均	
200	SGA	841.60	821.49	834.89	>5
	GQA	854.57	841.08	847.27	<3
	AEBGA	890.50	882.87	886.08	<3
500	SGA	1 974.75	1 952.24	1 968.57	>15
	GQA	2 016.24	2 001.55	2 010.45	<11
	AEBGA	2 074.15	2 049.88	2 062.28	<9
1 000	SGA	4 096.84	4 053.85	4 076.03	>26
	GQA	4 132.91	4 098.63	4 120.38	<17
	AEBGA	4 169.03	4 126.09	4 158.50	<16



(a) 200 个物品



(b) 500 个物品



(c) 1 000 个物品

图 3 3 种算法的实验结果比较

4 结束语

本文将定向变异机制引入到遗传算法中, 提出了基于主动进化遗传算法的 0-1 背包问题求解算法。实验结果表明, 该算法在全局寻优能力和执行效率上明显优于目前基于标准遗传算法和遗传量子算法的 0-1 背包问题求解方案。而且由于主动进化遗传算法其独特的编码和进化方案, 使得该算法思想非常适合推广到对其它复杂优化问题的求解过程中。

参考文献

- 1 陈国良, 王煦法, 庄镇泉, 等. 遗传算法及其应用[M]. 北京: 人民邮电出版社, 1996.
- 2 王晓东. 算法设计与分析[M]. 北京: 清华大学出版社, 2003.
- 3 Salomon R. Improving the Performance of Genetic Algorithms through Derandomization[J]. Software Concepts and Tools, 1997, 18(4): 175-184.
- 4 Han K H, Kim J H. Genetic Quantum Algorithm and Its Application to Combinatorial Optimization Problem[C]//Proceedings of the 2000 IEEE Congress on Evolutionary Computation. 2000: 1354-1360.
- 5 Cairns. The Origin of Mutation[J]. Nature, 1988, 3(35): 142-145.
- 6 史亮, 李海鹰, 杨俊安, 等. 基于主动进化的遗传算法研究[J]. 小型微型计算机系统, 2004, 25(5): 790-793.
- 7 史亮, 邹谊, 庄镇泉. 基于主动进化遗传算法的模糊聚类技术[J]. 小型微型计算机系统, 2005, 26(2): 204-208.