

遗传算法在含偏好分配问题上的应用

吴树猛

(中国矿业大学理学院, 徐州, 221116)

摘要: 含偏好分配问题是指, 每个人从一组项目中进行选择, 同时每个人都对不同的项目有不同的偏好。用遗传算法来解决这类分配问题, 其优点是算法能够给出一组满意解。采用平方函数来评价个体的适应性, 结果表明, 适应函数可以有效区分不同个体。

关键词: 运筹学; 遗传算法; 含偏好分配问题

中图分类号: O221

The application of genetic algorithm to the assignment problem with preferences

WU Shumeng

(Science School, China University of Mining and Technology, Xu Zhou, 221116)

Abstract: Assignment problems with preferences are defined as that every people make choices from a project list and they have different preferences to different projects. In this paper, a genetic algorithm is proposed to solve this problem. The advantage of this method is that a group of satisfactory solutions will be given. A square function is present in this paper to act as the fitness function, which is verified by the results to be very effective in choosing the chromosome with higher fitness.

Key words: operational research; genetic algorithm; assignment problem with preferences

0 引言

在很多情况下会遇到含偏好分配问题, 比如给学生安排宿舍, 或者在一些人数较多的组织中的排班问题等等。在这类问题中, 参加选择的人首先指出他们的偏好, 这可以用一个偏好矩阵表示(图1), 这个矩阵代表一个评分系统, 在矩阵中, “1”代表第一选择, “2”代表第二选择等等。当人数和项目数变得很大时, 就会发生安排上的矛盾, 这时寻求一个适合所有人的安排方案就会很困难。比如, 如果两个人选择同一个项目作为他们的第一选择, 这时只能把该项目分配给其中的一个人, 而给另一个人分配其他方案, 比如她/他的第二选择方案。这个第二选择方案又有可能是其他人的第一选择, 矛盾再次发生。此问题可在整数规划的框架内寻求解决^[1], 但是决策者通常只能得到一个解决方案。本文引入遗传算法的一个优点是, 对任意一个偏好矩阵, 算法都给出若干个安排方案, 这使得决策者可以比较不同方案的优缺点, 这是遗传算法的一个特点^[2,3,4]。

		项目						
		1	2	3	4	5	Λ	L
人 员	1	B	1	B	3	B	B	2
	2	2	B	B	1	B	B	3
	3	4	2	B	3	1	B	B
	4	B	1	B	B	2	3	B
	M	Λ			Λ		M	Λ
	T	3	B	B	1	B	B	2

作者简介: 吴树猛(1982-), 男, 硕士在读, 主要研究方向: 运筹学, 图论. E-mail: wushumeng@gmail.com

图 1: 偏好矩阵

Fig.1 Preferences matrix

1 数学模型

用 $T = \{1, 2, \dots, t\}$ 表示人员集合, $L = \{1, 2, \dots, l\}$ 表示项目集合 ($l \geq t$)。对于 $i \in T, j \in L$, 定义 $d_{i,j}$ 表示第 i 个人对第 j 个项目的偏好。偏好矩阵 ($t \times l$) 中的每个整数值都在集合 L 中, $d_{i,j} = 1$ 表示该项目为首选, $d_{i,j} = 2$ 表示该项目为第二选择, 以此类推。如果 $d_{i,j}$ 没有被赋予一个集合 L 中的整数值 (也就是人员 i 没有将项目 j 列入其偏好列表), 那么 $d_{i,j}$ 被赋予一个惩罚值 B (B 适当大)。这样, 该问题的数学模型为:

目标:

$$\min \sum_{i=1}^t \sum_{j=1}^l d_{i,j} \lambda_{i,j} \tag{1}$$

约束条件:

$$\sum_{j \in L} \lambda_{i,j} = 1 \quad \forall i \in T, \tag{2}$$

$$\sum_{i \in T} \lambda_{i,j} \leq 1 \quad \forall j \in L, \tag{3}$$

$$\lambda_{i,j} \in \{0, 1\} \quad \forall i \in T, j \in L, \tag{4}$$

这里:

$$\lambda_{i,j} = \begin{cases} 1, & \text{如果项目 } j \text{ 被安排给人员 } i \\ 0, & \text{其它情况} \end{cases}$$

且

$$d_{i,j} = B, \quad \forall i \in T, j \in L, \quad \text{如果 } d_{i,j} \notin \{1, 2, \dots, l\}.$$

等式 (2) 表明每个人能且只能被安排到一个项目, 而 (3) 式表明, 任一项目最多安排给一个人。因为小的偏好值有更大的吸引力 (偏好值为 1 表示该项目是首选), 因此目标函数要求最小化。

2 算法与求解

本文用由整数组成的规则结构 (t 维向量) 表示方案空间。每个染色体 (或向量) 含有 t 个基因, 每个基因用一个整数 j 表示, 这个整数表示第 j 个项目被分配给了第 i 个学生 (图 2)。这种表示使得模型的约束条件 (2) (3) 得到满足, 所以算法的任务就是如何使得目标函数取值最小。

人员	1	2	3	4	5	...	$t-1$	t
项目	3	5	8	6	1	...	l	9

图 2: 染色体构成

Fig. 2 Construction of chromosomes

2.1 初始化染色体群体

首先, 检查偏好矩阵去掉那些无须调整的项目安排, 如果一个人选择了某个项目作为他的第一选择, 而这个项目又不在其他任何人的偏好列表中, 那么该项目就是不需要调整的。对于剩余的人, 产生一个由 N 个染色体组成的集合, 染色体上的每个基因都是随机从每个人的偏好列表中随机取值。定义 $r_{i,g}$ 表示在 $g(g = 1, \Lambda, N)$ 染色体上分配给第 $i(i \in T)$ 个人的项目。

2.2 适应函数

染色体的适应值由个人根据他们的偏好列表对方案产生的“满意度”来衡量。偏好值按线性变化从 1 到最大可能 l (最大偏好值小于或等于 l) , 用下面的公式去计算染色体 g 的适应值 f_g :

$$f_g = \sum_{i \in T} q(d_{i,r_{i,g}}) \quad (5)$$

$q(\cdot)$ 定义为一个关于偏好的平方函数, 在含偏好分配问题中, 完全线性的适应函数是不合理的, 给一个人安排第二选择而不是第一选择引起的失望程度和给他安排第五选择而不是第四选择引起的失望程度并不是线性变化的。所以本文选择平方函数作为函数 $q(\cdot)$ 的定义, 将惩罚值 $\{1, 4, 9, \Lambda, l^2\}$ 分别分配给偏好 $\{1, 2, 3, \Lambda, l\}$ 。同时, 算法寻找那些使得适应函数取值较低的个体。当计算每个染色体的适应值的时候, 一个没有处在个人偏好列表中的项目 $r_{i,g}$ 将会得到一个适当大的惩罚值 B , 这个值被赋予 $d_{i,r_{i,g}}$ 。

2.3 运用联赛选择方法选择父代

用联赛选择方法从群体中选择父代。随机从群体中选择两个染色体, 保留最适应的那一个 (使得适应函数取值较小的) 染色体。这个最适应的染色体作为一个父代, 然后去用同样方法寻找另一个。这两个父代个体随后产生子代。

2.4 运用交叉算子产生子代

将交叉算子作用于两个父代个体产生新的子代。交叉算子通过在染色体上随机选择一个或两个交叉点然后交换染色体上的部分基因来形成新的子代。

如果一个项目被安排给两个或更多人, 这样的方案属于非可行解, 为了避免这样的方案进入群体, 我们对交叉算子做一点修改, 具体如下:

用 f_u 和 f_v 分别表示两个父代个体 U 和 V 的适应值, 用 C 表示这两个父代将要产生的子代染色体。用下标 i 表示染色体上的第 i 个基因 ($i = 1, \Lambda, t$)。

如果 U_i 、 V_i 和 C_k , $k = 1, \Lambda, i-1$ 都不同, 那么:

- ①. 如果 $U_i = V_i$, 则令 $C_i = U_i = V_i$;
- ②. 如果 $U_i \neq V_i$, 那么
 - (a) C_i 以概率 $f_v / (f_u + f_v)$ 取值 U_i
 - (b) C_i 以概率 $f_u / (f_u + f_v)$ 取值 V_i

其次, 如果 U_i 和 V_i 都已经是基因 C_k , $k = 1, \Lambda, i-1$ 中的某一个, 则随机取一个与 C_k ,

$k = 1, \Lambda, i-1$ 不同的项目, 然后将它分配给 C_i 。

其他情况下, 对于 U_i 和 V_i , 将不在 $C_k, k = 1, \Lambda, i-1$ 中的那一个分配给 C_i 。子代 C 的每一个基因都通过这个过程产生。这个过程在保证解的可行性的同时, 以较大的概率保留了父代中的具有更高适应值的个体的适应性。

2.5 变异

子代的变异和交叉算子一样给遗传算法提供了随机性的因素。如果没有变异, 遗传算法会很快收敛到一个局部极小值, 因为群体缺乏足够的基因多样性。变异算子使得子代染色体 C 上的每个基因都会以给定的变异概率改变为它的等位基因。变异概率可选择若干不同的值。同样的, 为了保证解的可行性, 我们选择不在 C 中的基因为新的等位基因。

2.6 精英保留

如果当代最佳个体的适应值小于上一代最佳个体的适应值, 则用上一代群体中适应值前 10% 的个体替换当代群体中适应值后 10% 的个体。

2.7 算例结果

偏好矩阵取一个包括 30 个人和 40 个项目 (项目数大于或等于人数) 的矩阵。群体中含有 50 个染色体, 这符合 30-100 的数量要求^[5], 在 MATLAB 中编程求解^[6]。结果表明, 平方函数能够很好的区分不同可行解, 而且可以使算法更快收敛。图 3 表明了在不同变异概率 0.05 和 0.1 下得到的结果。在初始阶段, 适应值收敛很快, 较高的变异概率使得算法收敛的更快。

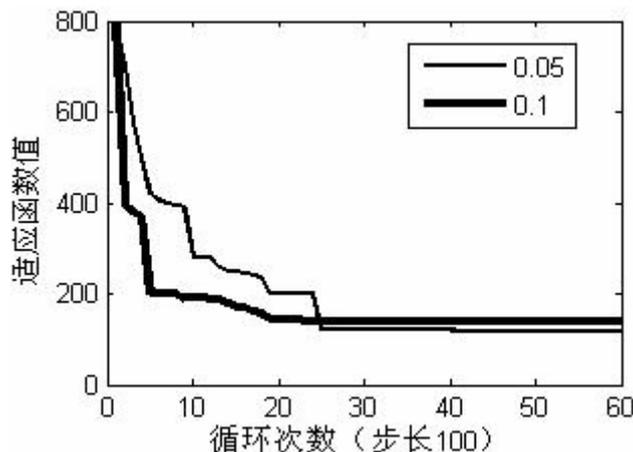


图 3: 在不同变异概率下的结果

Fig. 3 Results in different mutation rates

3 结论

本文引入遗传算法辅助解决含偏好分配问题, 它的一个突出优点是, 可以为决策者提供一组满意解。一个线性评价函数在此问题中是不合理的, 算法结果显示, 采用平方函数作为惩罚函数是很有效的。

[参考文献] (References)

- [1] 甘应爱, 田丰, 李维铮, 等. 运筹学[M]. 北京: 清华大学出版社, 2005.
- [2] 文军. 机场停机位分配问题的遗传算法[J]. 科学技术与工程, 2010, 10(1): 135—139.
- [3] 梁存利, 郑树团. 货架分配问题的一个有效混合遗传算法[J]. 计算机工程与应用, 2008, 44(8): 189—192.
- [4] 聂明泓, 杨丽英, 聂义. 任务分配问题的建模与求解[J]. 小型微型计算机系统, 2009, 30(4): 710—715.
- [5] 李敏强, 寇纪淞, 林丹, 等. 遗传算法的基本理论与应用[M]. 北京: 科学出版社, 2002.
- [6] 蔡旭晖, 刘卫国, 蔡立燕. MATLAB 基础与应用教程[M]. 北京: 人民邮电出版社, 2009.