

## 基于遗传算法的点群选取模型技术

武占峰, 吴裕树

(北京理工大学 计算机科学与工程系, 北京 100081)

(jonfond@126.com)

**摘要:**在地图上指定区域内依据一定的约束条件选取一组目标点(称为点群),很多领域都会用到。结合点群目标选取的约束条件和遗传算法的基本原理与特点,设计了一种基于遗传算法的点群目标选取模型。考虑到要最大限度地保持点群的多样性、在内部各地段的分布密度等因素,使用一种家族内相关选择的方法,并提出非固定基因位交叉变异的改进策略。实际计算表明,该算法性能稳定、搜索效率高,节省时间和空间,能有效地避免算法的“早熟”现象,且快速找到全局最优解。

**关键词:**点群;遗传算法;交叉;变异;早熟

**中图分类号:**TP181 **文献标识码:**A

## Model of points selection in map using genetic algorithms

WU Zhan-feng, WU Yu-shu

(Department of Computer Science & Engineering, Beijing Institute of Technology, Beijing 100081, China)

**Abstract:** Points selection in map is used in many fields. Combining the basic principle and characteristics of genetic algorithms with the basic principles of points selection, this paper designed a model of points selection in map based on genetic algorithms. In order to keep the variety of the points, and the internal distribution and density farthest, this paper proposed a new method with Elitist Recombination which employed an improved strategy of the ambulatory loca crossovers and mutations. Experimental results show that the method is timesaving and space-saving, it can avoid prematurity and quickly find optimal solution with stable performance and high search efficiency.

**Key words:** points; genetic algorithm; crossover; mutation; premature

## 0 引言

随着地图制图及相关学科的发展,数字地图自动综合的研究越来越受到重视。点群目标作为地图上的基本要素不仅大量存在,而且同线状要素一样是人们进行研究的重点。很多领域,例如决策支持、网络规划、气象学等领域,都需要在地图上指定区域内依据一定的约束条件选取一组目标点,根据各自具体的要求,定出选取点群的约束条件,然后使用合适的算法进行点群选取。

遗传算法(Genetic Algorithm, GA)是一种基于生物进化论和自然遗传学说的自适应随机全局优化算法。遗传算法的操作对象是种群,这个群体被限制在问题特定的环境里,根据预定的目标函数对每个个体进行评价,给出一个适应度值。从初始种群出发,在当前种群中选择个体,“适者生存”,使用交叉和变异来产生下一代种群。后代继承了父代的一些优良性状,“个体”的群体将逐步朝着更优解的方向进化。如此一代代演化下去,直到满足期望的终止条件。

遗传算法利用了生物进化和遗传的思想,具有许多与传统优化算法不同的特点:

1) 采用编码的方式,直接处理的对象是参数的编码集而不是问题参数本身,通过优良染色体基因的重组,遗传算法可以有效地处理传统上非常复杂的优化函数求解问题,并且具有良好的可操作性与简单性;

2) 直接以目标函数值作为搜索信息,而无需导数、梯度

等高价值信息,因而适用于大规模、高度非线性的不连续多峰函数的优化,以及不连续、不可微甚至无解析表达式的目标函数的优化,能以很大的概率求得全局最优解<sup>[1]</sup>,具有很强的通用性;

3) 同时使用多个搜索点的搜索信息,操作对象是一组可行解,而非单个可行解,搜索轨道有多条,而非单条,隐含并行性算法机制;

4) 具有很强的鲁棒性,即在存在噪声的情况下,对同一问题的多次求解中得到的结果是相似的<sup>[2,3]</sup>。

遗传算法提供了一种求解复杂系统优化问题的通用框架<sup>[3]</sup>,它不依赖于问题的具体领域,对问题的种类有很强的鲁棒性。

## 1 用于点群选取的遗传算法设计

遗传算法作为一种求解问题的高效并行全局搜索方法,越来越受到普遍重视,文献中提出了很多遗传算法的有效应用<sup>[2,3]</sup>,与各参考文献中的遗传算法相比,本文使用遗传算法在点群选取问题上设计出了新的算子和算法。

### 1.1 遗传编码的设计

编码问题实际是从问题空间到表达空间的映射问题,本文编码方案在网格数据模型的基础上考虑到地理信息的特殊性进行设计。网格数据模型的实质是将地理信息离散化,按照符合实际情况的跨度将待分析的矩形区域划分为网格,将 $x, y$ 坐标放在一起,采用二进制编码。二值编码具有简单易

行,符合最小字符集编码原则和便于用模式定理进行分析等几个优点。

对于由  $m \times n$  个网格组成的矩形,其编码长度为  $l = \log_2(m \cdot n) - 1$ 。

## 1.2 遗传算子的设计

遗传算法包含选择、交叉和变异三个基本操作,对于遗传算法中的选择算子、交叉算子和变异算子,其操作须事先确定选择策略、交叉率和变异率。

为达到更好的搜索效率,依据地理信息数据的特点使用启发式遗传算法,根据种群进化情况,本算法动态地调整遗传算子,维持了种群的多样性,克服过早收敛,并加快了搜索速度。

### 1) 选择算子和交叉算子

选择即从群体中选择优秀个体、淘汰劣质个体的操作。交叉指把父代个体的部分结构加以替换重组生成新个体的操作。在点群选取的搜索过程中可以从种群中按照给定的概率依次选择要进行交叉操作的个体,交叉操作采用两点交叉的方式,使用父子混合选择(father-offspring combined selection)算子来选择群体中要被保留的个体。

父子混合选择算子的基础是 ER (Elitist Recombination) 法,这种方法以家庭为单位来进行选择、交叉操作。家庭由两个亲个体以及这两个亲个体交叉生成的两个子个体组成的四个点中进行选择<sup>[2]</sup>,选择最大的两个个体一起遗传至下一代,其他两个个体则删除。对群体中的所有家庭都按同样的方式进行处理。

这种父子混合选择的方法在解空间的搜索过程中具有良好的收敛性,并且不影响解探索的效率;同时还有利于减少对基准点的评价次数,得到更好的次优解。

### 2) 变异算子

变异是对群体中的个体串的某些基因座上的基因值作变动,在地域选取中由于采用二进制编码,所以直接将要进行变异的位取反即可。为了维持种群多样性,采用非一致变异的策略,在遗传过程中动态地调节变异率。

### 3) 基因位选择

交叉变异操作的对象是基因位,采用动态循环决定基因位,提供给交叉变异操作的方法。

在二进制编码下,任取两个父代  $X_1 = x_1x_2 \cdots x_n$  和  $Y_1 = y_1y_2 \cdots y_n$ ,其中对于  $i = 0, 1, 2, \cdots, n, x_i$  和  $y_i$  为相应的基因位,对第  $i$  个基因位进行单点交叉操作,则可以产生两个子代  $X_2 = x_1x_2 \cdots x_{i-1}y_ix_{i+1} \cdots x_n$  和  $Y_2 = y_1y_2 \cdots y_{i-1}x_iy_{i+1} \cdots y_n$ 。可以发现,只要  $x_i \neq y_i$ ,则单点交叉变异后子代与父代的距离随着  $i$  的不同而不同。对于  $i = 1, 2, \cdots, n, i$  越小则子代与父代的距离越大,适应度值的变化可能很大;而  $i$  越大则距离越小,适应度值的变化可能很小。

执行遗传算法所处的代数  $g$  对编码长度  $L$  求余,所得基因位  $i (= g/L)$  即提供给交叉变异操作,每个循环(不妨设  $g > L$ ) 中,交叉变异操作后,其父子点之间适应度函数值的变化一般是由大到小,符合求取全局最优解的思想。

在传统遗传算法中按照一定概率进行交叉变异,单纯将新产生的点加入下一代。在本文所设计的算法中,使用了父子混合选择,对交叉变异后的点比较其适应度选优进入下一代,这样做避免了优良个体的流失,但有可能造成“早熟”。作为弥补,为了较好地保持群体内点的多样性,设计了一种方法,使交叉变异操作在不同遗传代数中,在循环动态选择的基因位上进

行,避免造成成熟前收敛的“早熟”而产生局部最优解的现象,在大小不同的距离上,均匀增加群体内点的多样性。

## 1.3 评价函数的设定

适应度函数采用加权的方法,单个因素突出不一定能被选中,只有各个需要考虑的因素都较优的点才有较高的适应度。在适应度函数中,若需要重点或者优先考虑某些因素,可以相应增加该因素在计算适应度时的权重。

在适应度的计算过程中,有的情况下可以直接认为适应度值为零而无需使用适应度函数进行计算,以提高算法的执行效率。适应度取零包括如下几种情况:

1) 对于矩形网格区域内的网格点不在用户指定区域中的情况;

2) 对不能找到相应网格点的基因编码其适应度也直接设为零;(因为遗传基因的编码为二进制编码,当编码长度为  $l$  时,其整个编码的空间大小为  $2^l$ ,而此时的矩形网格区域的所有网格点可能并不能与之完全对应,这种冗余在算法中是允许的)

3) 当选取点完全不符合约束条件时,将确定要舍弃的网格点对应的个体适应度也设置为零。

## 2 算法执行的具体步骤

在程序中,遗传算法相关部分执行的具体步骤如下:

1) 在地图指定区域中的点群自动选取,初始种群应该尽量均匀地分布在整个矩形网格区域中,因此用随机方法直接在整个搜索空间中生成一组采样点,作为初始种群。建立网格索引,将初始选取点的  $x, y$  网格坐标索引综合起来编码;

2) 对每个候选的点,使用上面设计的适应度算法计算其适应度,保存起来,以备查询;

3) 从候选点的集合中,选出一个适应度较高点的集合。这些点直接作为下一代的成员(精英策略,Elitist Strategy),置于下一代种群的一部分;

4) 下一代另一部分点的产生方式是:以上一代点的集合为基础,进行交叉变异,然后对交叉变异后的点比较其适应度,选取高者进入下一代,从而产生完整的新种群,如果算法没有达到结束的条件,新种群就作为下一代选取的候选点;

5) 重复 2)~4) 不断优选,直到所设定的遗传代数,其结果不保证覆盖所有可行解,但收敛于最优;

6) 根据指定的约束条件,将所得到的结果点群排序,在地图上显示出来。

## 3 算法结果及分析

### 3.1 算法执行结果

设置不同的遗传代数,分别执行程序 5 次,记录选取出的符合约束条件的点数和消耗的时间,求其平均值,列于表 1。

表 1 不同的遗传代数下点群选取结果

遗传代数	平均结果 点数	耗时/s	遗传代数	平均结果 点数	耗时/s
5	22.4	4.015	70	63.6	5.487
10	30.4	4.088	80	63.4	5.692
20	36.6	4.263	90	64.4	5.804
30	43.2	4.590	100	65.0	6.106
40	50.8	4.771	150	66.6	6.746
50	55.6	4.922	200	67.2	7.008
60	60.2	5.153			

由表 1 可以发现:随着遗传代数增加,选取出来的结果点数也呈增加趋势,遗传代数增加到一定程度(本统计结果为 80 代左右),结果点数的增加逐渐减缓甚至停滞。折中考虑程序运行消耗的时间和结果要求,应用中选定的遗传代数为 100。

本文选取穷举算法作为本算法结果的对比参考。地图上点群选取有很多方法,穷举算法是其中构思较为简单的一种,即将所指定的区域内所有点按照约束条件排序,求出最优解。该方法因为数据量庞大、计算时间较长而难以满足实时系统的要求,但是作为一种验证实验结果的参考却十分可靠。使用穷举算法在等同条件的选取中得到的点数是 71。本文所设计的算法接近最优解,在本项目中可以满足需求。

### 3.2 算法分析

本算法开始是进行初始点群的选取,得出  $n$  个点。初始化阶段对点群中的每个网格点进行数据模型构建和映射,读取或者求取各个网格点的属性值。为提高效率,可以只对初始点群各个点的属性进行一次计算,保存起来。设构建每个网格所用的时间为  $t_m$ ,则初始化的时间为:

$$f_{ini}(t) = t_m \times n$$

$t_m, n$  皆为常数,因此,数据初始化阶段的时间复杂度应该为  $O(1)$ 。

设遗传算法的群体大小为  $ps$ ,最大更新次数为  $mg$ ,遗传算法操作所用时间为  $t_a$ ,则全部执行时间为:

$$f_{ga}(t) = mg \times ps \times t_a$$

$mg, ps$  皆为常数,所以执行时间的时间复杂度为  $O(1)$ 。

则整个过程的时间为:

$$f_{total}(t) = f_{ini}(t) + f_{ga}(t) = t_m \times n + mg \times ps \times t_a$$

可见只用常数时间即可以实现点群选取,与指定区域的大小、网格数目无关。

空间复杂度即为所设定的种群大小  $N$  的 2 倍(用于同时存放新、旧点群),是常数  $O(1)$ 。

穷举算法需要首先获取所有网格点的参数,然后按照关心的参数排序,即使采用时间复杂度最小的堆排序算法,其时间复杂度为  $O(n \log n)$ ,大于  $O(1)$ 。空间复杂度是指定区域内的全体点,设矩形网格区域长、宽的网格数目为  $a$  和  $b$ ,则初始种群中网格的数目为  $n = a \times b$ ,所需空间随着指定区域的增大而增大,大于  $O(1)$ 。

由以上分析、比较可知,本文所设计的算法在求解地图上点群选取问题中,在保证结果质量的前提下,具有节省空间、快速、高效等特点。

### 参考文献:

- [1] 高汉平,康立山,杨族桥,等. 基于自适应杂交、变异率的演化算法[J]. 黄冈师范学院学报, 2003, 23(3): 57.
- [2] 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.
- [3] 玄光南,程润伟. 遗传算法与工程设计[M]. 北京: 科学出版社, 2000.
- [4] 毕思飞. 基于遗传算法的决策支持技术研究[D]. 北京理工大学, 2004.
- [5] 邓红艳,武芳,钱海忠. 基于遗传算法的点群目标选取模型[J]. 中国图像图形学报, 2003, 8(A)(8): 970.

(上接第 614 页)

在调度问题中,作业可看成是事件,作业之间的定性关系就是时间区间间的关系。调度过程就是确定作业的开始和结束时间,即作业所在时间区间的开始和结束时间的过程。

例 1 调度问题:有  $A, B$  两台机器资源,4 个如下的作业:

$$J_1: e_1 = \langle ti_1, tl_1 \rangle = \langle \langle is_1, if_1 \rangle, \langle 5, 9 \rangle \rangle$$

$$J_2: e_2 = \langle ti_2, tl_2 \rangle = \langle \langle is_2, if_2 \rangle, \langle 6, 13 \rangle \rangle$$

$$J_3: e_3 = \langle ti_3, tl_3 \rangle = \langle \langle is_3, if_3 \rangle, \langle 8, 11 \rangle \rangle$$

$$J_4: e_4 = \langle ti_4, tl_4 \rangle = \langle \langle is_4, if_4 \rangle, \langle 7, 10 \rangle \rangle$$

并且有以下的条件:  $dp(\langle is_2, if_1 \rangle) = \langle 1, 7 \rangle$ ,  $di(\langle ti_4, ti_3 \rangle) = \langle dsf, \langle 0, 5 \rangle, dss, dff \rangle$ 。

解 由  $dp(\langle is_2, if_1 \rangle)$  的值均为正数可知,  $ti_1$  a-b  $ti_2$ , 因此机器  $A, B$  分别从确定时间点 0 分别开始运行作业  $J_1$  和  $J_3$ , 并将确定时间点 0 作为  $J_1$  和  $J_3$  的开始时间,又已知  $J_1$  和  $J_3$  的时间长度  $tl_1$  和  $tl_3$ , 由 CP 关系可确定  $if_1$  和  $if_3$  的分别为  $\langle 5, 9 \rangle$  和  $\langle 8, 11 \rangle$ 。根据  $dp(\langle is_2, if_1 \rangle) = \langle 1, 7 \rangle$  可推知  $is_2 = \langle 10, 12 \rangle$ , 再由  $tl_2 = \langle 6, 13 \rangle$ , 得知  $if_2 = \langle 18, 23 \rangle$ , 同理可以由条件  $di(\langle ti_4, ti_3 \rangle)$  推算  $J_4$  的开始和结束时间。最后的结果如下:

$$J_1: e_1 = \langle \langle \langle 0, 0 \rangle, \langle 5, 9 \rangle \rangle, \langle 5, 9 \rangle \rangle \quad A$$

$$J_2: e_2 = \langle \langle \langle 10, 12 \rangle, \langle 18, 23 \rangle \rangle, \langle 6, 13 \rangle \rangle \quad A$$

$$J_3: e_3 = \langle \langle \langle 0, 0 \rangle, \langle 8, 11 \rangle \rangle, \langle 8, 11 \rangle \rangle \quad B$$

$$J_4: e_4 = \langle \langle \langle 11, 13 \rangle, \langle 20, 21 \rangle \rangle, \langle 7, 10 \rangle \rangle \quad B$$

## 5 结语

针对人工智能不确定时态信息的表达和推理问题,本文

提出了一个描述不确定时态信息的统一模型,该模型不仅可以描述不确定时态信息,还可以将确定时态作为不确定时态的一种特殊情况加以描述。模型在定义了各种时态表示之后,也定义时态对象之间的定性和定量关系,并探讨了关系之间的传递性问题。

在本文的例子中,通过模型描述事件和事件之间的关系,并进行了不确定时态信息的推理。从例子可以看出,不仅现实生活的各类和时间有关的事件可以用该模型描述,而且像计算机的调度问题等方面的不确定时态推理,也可以利用该模型来描述。

### 参考文献:

- [1] PANI AK, BHATTACHARJEE GP. Temporal representation and reasoning in artificial intelligence: a review [J]. Mathematical and Computer Modelling, 2001, 34(1): 55-80.
- [2] ALLEN JF. Maintaining knowledge about temporal intervals [J]. Communications of the ACM, 1983, 26(11): 832-843.
- [3] MCDERMOTT DV. A temporal logic for reasoning about processes and plans [J]. Cognitive Science, 1982, 6(2): 101-155.
- [4] DECHTER R, MEIRI I, PEARL J. Temporal constraint networks [J]. Artificial Intelligence, 1991, 49(1): 61-95.
- [5] 张师超. 基于间断区间的时态知识表示 [J]. 软件学报, 1994, 5(6): 14-18.
- [6] 王家殿. 具有不确定性的时态模型 [J]. 计算机科学, 1995, 22(6): 49-54.
- [7] 方思行, 陆颖. 一种可视化的时态推理方法 [J]. 暨南大学学报 (自然科学版), 1999, 20(3): 48-52.