随机化均匀设计混合遗传算法求解图的二划分问题

周本达[1],陈明华[2]
(1.皖西学院 数理系,安徽 六安 237012; 2.皖西学院 计算机科学与技术系,安徽 六安 237012)

摘 要:图的二划分问题是一个典型的 NP-hard组合优化问题,在许多领域都有重要应用.近年来,传统遗传算法等各种智能优化方法被引入到该问题的求解中来,但效果不理想.基于理想浓度模型的机理分析,利用随机化均匀设计抽样的理论和方法,对遗传算法中的交叉操作进行了重新设计,并在分析图的二划分问题特点的基础上,结合局部搜索策略,给出了一个解决图的二划分问题的新的遗传算法.通过将该算法与简单遗传算法和佳点集遗传算法进行求解图的二划分问题的仿真模拟比较,可以看出新的算法提高了求解的质量、速度和精度.

关键词:图的二划分;遗传算法;随机化均匀设计
中图分类号: TP18 **文献标识码**: A **文章编号**: 1673-4785(2009)01-0091-04

Solving the 2-way graph partitioning problem using a genetic algorithm based on randomized uniform design

ZHOU Ben-da[1], CHEN Ming-hua[2]
(1. Department of Mathematics and Physics, West Anhui University, Lu'an 237012, China; 2. Department of Computer Science and Technology, West Anhui University, Lu'an 237012, China)

Abstract: 2-way graph partitioning is a typical NP-hard combinatorial optimization problem, and has widespread applications in many domains. Many intelligent optimization methods, including traditional genetic algorithms, were recently introduced to solve this problem, but the results have not been ideal. Following analysis of the mechanisms of the ideal density model, the genetic algorithm (GA) crossover operation was redesigned using the principles and methods of randomized uniform sampling design. Furthermore, a new GA for solving 2-way graph partition was formulated. Simulations which compared results from this method with simple GA and Good Point GA for solving the 2-way graph partitioning problem showed that the new GA has superior speed, accuracy, and precision.
Keywords: 2-way graph partitioning; genetic algorithm; randomized uniform design

图的二划分在数字电路系统的芯片划分、网络管理,插件划分、分布式系统的数据流划分、并行系统的信息流划分等许多领域有着重要应用.众所周知,图的二划分问题是一个 NP-hard组合优化问题,因此精确求出最优解是不可能的,较为实际的方法是尽快地发现其近似最优解.对于图的二划分向题,已提出了许多启发式的搜索算法,其基本思想是寻找当前二划分子集中与其他各节点连接最少的节点,将其移动到另外的子集中,如此反复移动,直到可行集不能改进为止.这种方法求解质量相对较高,但对初始划分的依赖性很强,陷入局部最优解的可能性较大.近年来,各种智能优化方法[1-2]被引入到该问题的求解中来.

遗传算法[1]作为一种全局优化搜索算法,由于其本身所具有的全局收敛性和隐并行性,加之其简单易用、鲁棒性强,能够轻易地获得问题的全局最优解;且问题越复杂,它相对于其他算法的优越性越明显,故十分适合解决这类问题.但应用经典遗传算法求解图的划分问题时,求解时间长、成功率低.因而,有必要针对具体问题的特点,对经典的遗传算法加以改进.

文献[3]对遗传算法(genetic algorithm, GA)中的两个理论基石"模式定理"和"隐性并行性"进行了分析,指出 GA的本质是一个具有定向制导的随机搜索技术,其定向制导的原则是:导向以高适应度模式为祖先的"家族"方向.文献[4]根据此机理,利用数论中的佳点集理论和方法[5]设计了一个新的

收稿日期: 2008-05-12
基金项目:安徽省高校省级自然科学研究资助项目(KJ2007B152);安徽省教育厅自然科学研究资助项目(2005KJ222, 2006KJ046B);安徽省高校青年教师资助计划资助项目(2007jql180).
通信作者:周本达. E-mail: bendazhou@163.com.

交叉操作,提高了 GA的效率,这种算法称为佳点集遗传算法.但在佳点个数 n取定后,佳点集的选取是确定的,不带随机性.为了克服此不足,在充分分析文献[3]中理想浓度模型的基础上,基于随机化均匀设计理论对遗传算法中的交叉操作进行重新设计,并结合图的二划分问题的局部优化技术[6],提出了解决图的二划分问题的基于随机化均匀设计的混合遗传算法.为了表明算法的有效性,对图划分测试网站提供的标准测试算例进行了计算机仿真,结果表明新算法在收敛速度和最优解的质量上均优于简单遗传算法和佳点集遗传算法.

1 随机化均匀设计方法

随机化均匀设计过程如下[7]:

1)对于固定的 t和 $n\in N$,选取均匀设计的生成向量 $(n: h_1, h_2, \cdots, h_t)$;

2)从多项分布 $\begin{pmatrix} 0 & 1 & \cdots & n-1 \\ \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix}$ 中抽取 $t-1$个独立同分布样本 $\eta_1, \eta_2, \cdots, \eta_{t-1}$,并令 $\eta_t = 0$;

3)令 $X_k = (x_{k1}, x_{k2}, \cdots, x_{kt})$, $k = 1, 2, \cdots, n$.其中

$$x_{kj} = \left\{ \frac{k \cdot h_j + \eta}{n} \right\}, \quad k = 1, \cdots, n,\ j = 1, \cdots, t \qquad (1)$$

这里 $\{x\}$表示取 x的小数部分,称式(1)给出的点集

$$\chi_k = \{X_k = (x_{k1}, \cdots, x_{kt}),\ k = 1, \cdots, n\} \qquad (2)$$

为随机化均匀设计点集.

称这种方法为在 t维单位立方体 $C^t = [0, 1]^t$中选 n个点的随机化均匀设计(randomized uniform design, RUD)方法.

若记 $a_{kj} = \left\{ \frac{k \cdot h}{n} \right\}$, $k = 1, \cdots, n,\ j = 1, \cdots, t$,则称点集

$$A_n = \{a_k = (a_{k1}, \cdots, a_{kt}),\ k = 1, \cdots, n\} \qquad (3)$$

为均匀设计点集[7].

实际上,这种随机化均匀设计是通过对均匀设计进行模 1随机平移而得到的.这种平移总共有 n^t个,而每组样本点个数为 n,这样正好将全体 n^t个格子都以同等机会抽到,因此,它具有较好的搜索能力.而且文献[7]的定理 2.2证明了随机化均匀设计点集的偏差比佳点集小,并且它能随机地取到所有的格子,所以其搜索能力更强,故会有比佳点集更好的表现.

2 遗传算子

2.1 染色体编码

基于图的二划分问题,这里采用二进制编码.例如 $x = [0\ 1\ 0\ 1\ 1\ 1\ 0\ 1]$表示将顶点集合 $V = \{1, 2, 3, 4, 5, 6, 7, 8\}$划分为 $V_1 = \{1, 3, 7\}$和 $V_2 = \{2, 4, 5, 6, 8\}$两个子集.

2.2 适应度函数

根据图的二划分定义及划分原则,因为图中所有边的权值和是一个常量,求属于不同分块的顶点之间的边的权值之和的最小值问题,实际上也就是求同一分块内各顶点之间的边的权值之和的最大值问题,据此定义适应度函数如下:

$$f(x) = g(x) - r \cdot u \cdot g(x). \qquad (4)$$

其中: $g(x) = \sum_{1 \le i \le j \le n} w(v_i, v_j)$, v_i与 v_j同属于同个分块; $w(v_i, v_j)$表示顶点 v_i与 v_j间的权值,在简单无向图中定义为

$$w(v_i, v_j) = \begin{cases} 1 & v_i \text{与} v_j \text{间有边;} \\ 0 & \text{否则.} \end{cases}$$

式(4)中第 2项为惩罚函数, $0 < r < 1$为惩罚系数,它根据个体违反约束条件的程度而定, r越大,约束条件要求越严格,否则约束条件比较宽松; u为解是否为合法解的判定系数,可定义为

$$u = \begin{cases} 0 & x \text{为合法解;} \\ 1 & \text{否则.} \end{cases}$$

2.3 选择算子

采用轮盘赌选择算子.即将所有染色体的适应值之和看作一个轮盘,每个染色体根据其适应值的大小划分在轮盘中所占据的范围.然后旋转轮盘,当轮盘停下来时,指针所对应的染色体即被选中,完成一次选择.重复上述过程,直到选择到所需要的染色体个数为止.

2.4 基于随机化均匀设计的杂交算子

设在传统的 GA算法基础上,在进行过复制后,对池中的染色体随机选择两个 A_1、A_2进行随机化均匀设计交叉操作.一般情况下是令:

$$A_1 = (a_1^1, a_2^1, \cdots, a_l^1),$$

$$A_2 = (a_1^2, a_2^2, \cdots, a_l^2),$$

$$J = \{i \mid a_i^1 \neq a_i^2, 1 \le i \le l\}.$$

不妨设 A_1, A_2的前 t个分量不同,后 $l-t$个分量相同,令模式

$$H = \{(x_1, x_2, \cdots, x_l) \mid i \in J,\ x_i = *;\ i \notin J,\ x_i = a_i^1\}.$$

由 A_1、A_2进行交叉(不管是单点交叉或是多点交叉),其子孙必属于 H,于是要在"高适应度模式为祖先的家族方向"上搜索出更好的样本,就是要在 H中搜索出更好的样本.随机化均匀设计交叉操作就是要在 H上利用随机化均匀设计方法找出好样本来.

对图的二划分问题,由于码串 0101与 1010表示的划分相同,在这里对于两个染色体 A_1、A_2首先直接比较,记录下不同值的位置存于 J_1;然后将 A_2

各位取反,再同 A_1 进行比较,记录下不同值的位置存于 J_2,取 J_1、J_2 中长度小的为 J,不妨令对应的模式仍为 H. 不同值的位置构成一个 t维立方体,记为 H′,然后在 H′上进行随机化均匀设计抽样,即要在 t 维单位立方体 $C^t = [0, 1]^t$ 中进行选 n个点的随机化均匀设计交叉操作,具体如下:

1)对于固定的 t和 $n \in N$,选取均匀设计的生成向量 $(n: h_1, h_2, \cdots, h_t)$;

2)从多项分布 $\begin{Bmatrix} 0 & 1 & \cdots & n-1 \\ \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \end{Bmatrix}$ 中抽取 $t-1$个独立同分布样本 $\eta_1, \eta_2, \cdots, \eta_{t-1}$,并令 $\eta_t = 0$;

3)令 $X_k = (x_{k1}, x_{k2}, \cdots, x_{kt})$, $k = 1, 2, \cdots, n$,

$$x_{kj} = \left\{ \frac{k \cdot h_j + \eta}{n} \right\}, \quad k = 1, \cdots, n, \ j = 1, \cdots, t$$

令交叉后产生的 n个后代中第 k个染色体,$B^k = (b_1^k, b_2^k, \cdots, b_l^k)$,其中:

$$b_m^k = \begin{cases} a_m^1, & m \notin J, \\ \langle x_{kj} \rangle, & m = t_j \in J, \ 1 \le j \le t \end{cases} \tag{5}$$

式(5)中:$1 \le k \le n$, $1 \le m \le l$, $\langle a \rangle$表示:若 a的小数部分小于 0.5,则 $\langle a \rangle = 0$;否则 $\langle a \rangle = 1$.

这样在其"家族"中,产生了 n个后代(依次取 $k = 1, 2, \cdots, n$所得),取其中适应值最大者(或最大的几个),作为交叉后的后代. 上述交叉操作,称为随机化均匀设计交叉操作.

2.5　变异算子

变异操作为:取染色体 A,随机整数 i, $A = (a_1, a_2, \cdots, a_l)$ 变异成新的染色体 $B = (a_1, a_2, \cdots, a_{i-1}, b_i, a_{i+1}, \cdots, a_l)$,其中 $b_i = \overline{a_i}$.

3　局部搜索技术

顶点的适应值定义为:一个顶点 j的适应值 $\lambda_j = g_j / (g_j + b_j)$,其中:$g_j$为与顶点 j同一个子集中的顶点与 j连接的边的权值和数;b_j为与 j不同的另一个子集中的顶点与 j连接的边的权值和数. 如果 $g_j + b_j = 0$,那么设定 $\lambda_j = 1$. 则求解图的二划分问题的局部搜索算法如下:

1) 对于一个划分 $\langle V_1, V_2 \rangle$,其中:$V_1 \cup V_2 = V$, $V_1 \cap V_2 = \varnothing$ 令 $U_{best} = \langle V_1, V_2 \rangle$.

2) 计算 V中各个顶点的适应值.

3) 选择 V中适应值最小的顶点设为 m,再在不含 m的子集中随机地选取另一个顶点 n,交换 m和 n,改写 V_1 和 V_2. 若新得的划分好于 U_{best},则改写 U_{best}.

4) 若没达到设定的交换次数,则转 2),否则算法终止并返回 U_{best}.

局部搜索算法通过对当前的顶点的适应值的比较来选择交换,对于顶点的适应值的定义既考虑了顶点在同一个子集中连接的边数又考虑了顶点的度. 这样可以避免过早地陷入局部最优解,有利于在更大的范围内搜索全局最优解.

4　求解图的二划分问题的随机化均匀设计混合遗传算法

给定交叉概率 p_c 和突变概率 p_m 后,随机化均匀设计混合遗传算法(genetic algorithm based on randomized uniform design, RGA)如下:

1)每次进行遗传操作,以概率 $f_i / \sum f_i$ 复制 A_i,其中 f_i 是 A_i 的适应度值.

2)以概率 p_c 对其进行随机化均匀设计交叉操作(产生 n个后代,n为待定参数).

3)以概率 p_m 进行变异遗传操作.

4)对新产生的染色体进行局部搜索操作.

5)把经过上述操作后得到的染色体都放到染色体池中,对新得到的染色体,计算其适应度值. 若假定染色体的容量一定,当染色体的个体超过容量时,就将适应度小的染色体从池中删去(或按 $a\%$进行删除).

6)进行上述的遗传算法至第 T代(T是预先给定的常数),在算法执行过程中记录适应度最大的染色体,即为所求的染色体,再进行解码得到最优解.

5　实验结果及分析

为了表明算法的有效性,分别用简单遗传算法、佳点集遗传算法和随机化遗传算法在同样条件下,即在 P4 3.0G PC机器上,采用 Matlab7.0计算平台对国际标准数据库[8]中的算例(见表 1)进行仿真,结果如表 2所示.

表 1　算例数据

Table 1　Example data

问题	顶点数	边数目	顶点平均度	顶点最大度	顶点最小度
Queen5_5	25	160	6.40	16	12
David	87	406	4.67	82	1
Miles250	128	387	3.02	16	0
Queen10_10	100	1 470	14.70	35	27
Queen13_13	169	3 328	18.03	48	36
Queen15_15	225	5 180	23.02	56	42

表 2 SGA、GGA和 UGA的实验结果比较

Table 2 Experimental result of SGA, GGA and UGA

问题	算法	N_{best}	V_{avg}	σ	N_{avg}
Queen5_5	SGA	60	63.12	2.93	51.31
	GGA	60	60.68	1.85	11.48
	RGA	60	60.00	0.00	1.40
David	SGA	83	94.01	5.90	149.72
	GGA	82	91.08	4.77	62.63
	RGA	82	87.88	3.26	23.96
Miles250	SGA	14	35.20	9.94	138.19
	GGA	18	47.14	13.57	65.34
	RGA	4	10.30	4.20	20.89
Queen10_10	SGA	495	515.68	21.81	156.19
	GGA	500	515.64	2.92	47.90
	RGA	495	495.00	0.00	3.76
Queen13_13	SGA	1 188	1 279.20	42.09	174.73
	GGA	12 320	1 364.20	47.07	125.79
	RGA	1 092	1 092.30	1.71	18.74
Queen15_15	SGA	1 930	2 066.30	57.44	176.08
	GGA	2 148	2 264.80	32.44	123.58
	RGA	1 680	1 685.30	11.74	15.37

其中: SGA为简单遗传算法; GGA为佳点集遗传算法; RGA为基于随机化均匀设计的混合遗传算法. 算法参数为:群体规模为 100;交叉和变异概率: $P_c = 0.9$, $P_m = 0.05$;最大迭代代数为 200. 算法中惩罚系数:

$$r = 1 - \left(\frac{1}{2}\right) \mid \max(\mid V_1 \mid, \mid V_2 \mid) - n/2 \mid,$$

n为顶点个数; 当 $(\max|V_1|, |V_2|) - n/2) \leq 5$时 $u = 0$,否则 $u = 1$. 为避免遗传算法的随机性,对每个标准算例在同一台机器上进行连续 100次计算,记录如下结果:

1) 100次运行中求得的最好解 (记为 N_{best})和 100次运行所得解的平均值 (记为 V_{avg}); 2) 100次运行求得解的标准差 (记为 σ); 3) 100次运行中每次收敛时代数的平均值 (记为 N_{avg}).

由表 2可以得出:随机化均匀设计混合遗传算法每次得到的解均好于简单遗传算法和佳点集遗传算法;而且在解的平均值、平均收敛代数、标准差等指标上均好于简单和佳点集遗传算法. 由此说明随机化均匀设计混合遗传算法在搜索能力、收敛速度以及避免早熟等各项指标上均好于简单遗传算法和佳点集遗传算法.

6 结束语

文章以遗传算法的"理想浓度模型"为基础,充分分析其运行机理,利用统计抽样的理论和方法对算法中的交叉操作进行了重新设计. 分析图二划分问题的特点,结合局部搜索技术,提出了新的解决图二划分问题的混合遗传算法,克服了佳点集遗传算法中佳点选取不带随机性的缺点,实例仿真得出的结果表明了新算法的有效性和先进性. 今后的工作是进步分析该方法的深层次的数学基础以及该方法在组合优化问题中的有效性和先进性.

参考文献:

[1] KANG S J, MOON B R. A hybrid genetic algorithm for multi-way graph partitioning[C]//Proc Genetic & Evolutionary Computation Conf (GECCO-2000). San Francisco, USA: Morgan Kaufmann, 2000: 159-166.

[2] HENDRICKSON B, KOLDA T G. Graph partitioning models for parallel computing[J]. Parallel Compute, 2000, 26(12): 1519-1534.

[3] 张 铃,张 钹. 遗传算法机理的研究[J]. 软件学报, 2000, 11(7): 945-952.
ZHANG Ling, ZHANG Bo. Research on the mechanism of genetic algorithms[J]. Journal of Software, 2000, 11(7): 945-952.

[4] 张 铃,张 钹. 佳点集遗传算法[J]. 计算机学报, 2001, 24(9): 917-922.
ZHANG Ling, ZHANG Bo. Good point set based genetic algorithm[J]. Chinese Journal of Computers, 2001, 24(9): 917-922.

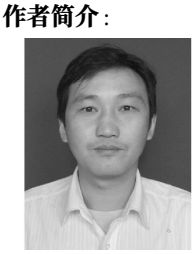
[5] 华罗庚, 王 元. 数论在近似分析中的应用[M]. 北京: 科学出版社, 1978.

[6] SOPER A J, WALSHAW C, CROSS M. A combined evolutionary search and multilevel optimization approach to graph partitioning[J]. Journal of Global Optimization, 2000, 29(2): 225-241.

[7] 王兆军, 张润楚. 均匀设计抽样的偏差[J]. 数学物理学报, 1997, 17(2): 207-217.
WANG Zhaojun, ZHANG Runchu. Descripency of uniform desingn sampling[J]. Acta Mathematica Scientia, 1997, 17(2): 207-217.

[8] WALSHAW C. The graph partitioning archive[EB/OL]. [2008-04-15]. http://staffweb.cms.gre.ac.uk/~wc06/partition/.

作者简介:



周本达,男, 1974生,副教授. 主要研究方向为遗传算法、多 Agent系统. 发表学术论文 10余篇.

陈明华,男, 1954生,教授,主要研究方向为遗传算法、统计建模中的大样本理论.