

文章编号: 1001-9081(2007)07-1806-03

基于遗传禁忌算法结合解决排课问题

陈守家¹, 付霞^{2,3}, 周欣³

(1. 山东潍坊商业学校 信息技术系, 山东 潍坊 261011; 2. 上海师范大学 计算机系, 上海 201804;
3. 渭南师范学院 计算机系, 陕西 渭南 714000)
(csjsj2002@163.com; zxf1978@126.com)

摘要:排课问题是一典型 NP-Hard 问题, 通常可以使用遗传算法进行解决, 把遗传算法与局部搜索方法禁忌算法有机结合起来, 是改进遗传算法性能的一个卓有成效的方法。使用遗传禁忌算法解决排课问题, 并且通过改变个体适应度的计算方法, 避免了排课中课表的两极分化现象。通过实验, 该方法可以取得较好的排课结果。

关键词:排课问题; NP-Hard 问题; 遗传算法; 禁忌算法; 遗传禁忌算法

中图分类号: TP18; TP311.52 **文献标志码:** A

Solution to course scheduling problems based on genetic algorithm and tabu search

CHEN Shou-jia, FU Xia^{2,3}, ZHOU Xin³

(1. Department of Information Technology Shandong Weifang Commercial School Weifang Shandong 261011 China;
2. Department of Computer Shanghai Normal University Shanghai 201804 China;
3. Department of Computer Weinan Teachers' College Weinan Shaanxi 714000 China)

Abstract Course scheduling problem is a type of NP-hard problems. Usually it can be solved by utilizing genetic algorithm. Tabu search algorithm was combined with genetic algorithm to improve the performance. Furthermore, by altering the individual fitness computation method, the curriculum schedule polarization phenomena can be avoided. Experimental results indicate that the method is feasible and has better performance.

Key words: course scheduling problem; NP-hard problem; genetic algorithm; tabu search; GATS algorithm

0 引言

排课问题实际是一时间表问题, 由于该问题具有较大实用价值和问题模型的代表性, 吸引了许多学者对该问题进行研究, 从各个角度对问题进行了分析, 提出了许多模型和算法。如: 王璐等利用多 Agent 之间的协商来实现排课^[1]; 潘以锋则通过对教学任务安排优先级, 采用优化资源查找算法来解决排课问题^[2]。但这些模型和算法有的智能性并不强, 需要过多的启发信息和人的参与, 有的虽然可以排出课程表, 但课表的质量并不高, 如课程对时间段的特殊需求未被考虑, 或者教师的偏好未被考虑等。事实上, 在 1975 年, 排课问题已被 S. Even 等人论证为 NP 完全类问题, 而对于这类问题, 只能利用智能算法求得满意解。遗传算法 (Genetic Algorithm, GA) 是一种适合求解那些带有多参数、多变量、多目标和在多区域但连通性较差的 NP-Hard 优化问题的智能优化算法, 而排课问题正具有这些特点。因此, Yen-Zen Wang^[3] 利用 GA 算法对该问题进行了研究, 并解决了排课问题。

然而由于 GA 本身的固有不足, 把 GA 与局部搜索方法有机结合起来, 是改进 GA 性能的一个卓有成效的方法。这种混合型遗传算法不但模拟了生物种群的学习过程, 而且事实

上还模拟了种群的个体在其生命周期内具有学习行为这一生物现象^[4]。禁忌算法 (Tabu Search, TS) 是局部搜索算法的一种改进方法, 因此将 GA 结合 TS 来求解排课问题, 是一种可行的方法。实验结果表明使用 GA 结合 TS 的方法完全可以解决排课问题, 通过与文献 [3] 所述方法进行比较, 本文算法所求得的课表比较优越。

1 遗传禁忌算法

GA 是 John H. Holland 根据生物进化的模型提出的一种优化算法, 它是基于进化过程中的信息遗传机制和优胜劣汰的自然选择原则和搜索算法。它从一个种群开始, 利用选择、交叉、变异等遗传算子对种群进行不断进化, 最后得到全局最优解或近似最优解。

TS 是一种亚启发式随机搜索算法, 是局部邻域搜索算法的推广。Glover 在 1986 年首次提出这一概念, 进而形成一套完整算法。它从一个初始可行解出发, 选择一系列的特定搜索方向 (移动) 作为试探, 选择实现让特定的目标函数值变化最多的移动。为了避免陷入局部最优解, TS 搜索中采用了一种灵活的“记忆”技术, 对已经进行的优化过程进行记录和选择, 指导下一步的搜索方向, 这就是 Tabu 表的建立。Tabu 表

收稿日期: 2007-01-17 修回日期: 2007-03-19 基金项目: 渭南师范学院科研基金资助项目 (04YK519)。

作者简介: 陈守家 (1976-), 男, 山东潍坊人, 讲师, 主要研究方向: 人机交互与普适计算; 付霞 (1978-), 女, 黑龙江人, 讲师, 硕士研究生, 主要研究方向: 计算机网络; 周欣 (1979-), 男, 湖南人, 讲师, 硕士研究生, 主要研究方向: 模式识别、智能计算。

中保存了最近若干次迭代过程中所实现的移动的反方向移动, 凡是处于 Tabu 表中的移动, 在当前迭代过程中是不允许实现的, 这样可以避免算法重新访问在最近若干次迭代过程中已经访问过的解群, 从而防止了循环, 帮助算法摆脱局部最优解。另外, 为了尽可能不错过产生最优解的“移动”, TS 搜索还采用“释放准则”的策略。TS 的搜索速度比 GA 快, 但 TS 对于初始解具有较强的依赖性。一个较好的初始解可使 TS 在解空间中搜索到更好的解, 而一个较差的初始解则会降低 TS 的收敛速度。为此, 人们往往使用启发式算法来获得一个较好的初始解, 提高 TS 的性能。TS 的另一缺陷是在搜索过程中初始解只能有一个, 在每代也只是把一个解移动到另一个解, 而不像 GA 那样每代都是对解集(群体)进行操作。

由于两种算法都存在着无法避免的缺点^[4-5], 因此提出基于这两种算法的一种混合策略来解决排课问题, 采用 TS 作为遗传算法 GA 的变异算子。这一策略将 TS 嵌入 GA 中, 运用 TS 强大的爬山能力来有效地避免 GA 存在的“早熟”现象。

2 问题描述及数学模型

排课问题主要解决的是如何为所有课程(包含教师信息)安排合适的时间段与合适的教室, 使课程不产生冲突, 即满足一些硬约束条件; 同时, 应该尽量考虑满足一些使教学效果提高的条件, 这些就是软约束条件。

2.1 排课问题的硬约束及数学模型

常见硬约束有: 一个老师不能在同一时段在两个教室同时授课; 一个教室不能在同一时段安排两次课; 每个班级在同一时间只能上一门课; 学生人数不能超过教室容量。

首先设以下变量, 集合及函数:

- 1) 设班级集合为 C 教师集合为 T 时段集合为 D
- 2) C_i 表示 班级的人数。
- 3) $\tau(i, j)$ 表示教师 教授课程 的时段; $\pi(t, i)$ 表示 时段 班级所在上课教室; $\varphi(t, i)$ 表示 时段 班级的上课课程; $v(i)$ 表示 教室的容量; $I_1(i)$ 表示教师 教授的课程集合; $I_2(i)$ 表示班级 的课程集合。

则我们的问题应该满足以下硬约束条件:

$$s.t. \left(\sum_{t \in T} \sum_{k \in I_1(i)} \tau(i, j) = \tau(i, k) \right) = 0 \tag{1}$$
$$\left(\sum_{t \in D} \sum_{j \neq i} \pi(t, j) = \pi(t, i) \right) = 0 \tag{2}$$
$$\left(\sum_{t \in D} \sum_{j \neq i} \varphi(t, j) = \varphi(t, i) \right) = 0 \tag{3}$$
$$\forall \pi(t, i) \geq C_i, \forall t \in T, i \in C \tag{4}$$

式(1)表示一位教师不能在同一时段安排两次课; 式(2)表示一个教室不能在同一时段安排两次课; 式(3)表示每个班级在同一时间只能上一门课; 式(4)表示学生人数不能超过教室容量。

2.2 排课问题的软约束

满足硬约束的课表是可使用的, 但却不一定是令人满意的。满意的课表还应满足一些软约束条件如: 同一门课在一周内的分布尽可能均匀; 某些特定的课最好不排特定的时段, 如自习课和体育课最好不排每天的一、二节; 尽量满足每位教师的偏好等。对于软约束, 根据每个条件的重要程度, 将定义一个相对应的权值, 然后参与适应度的计算, 从而影响排课结果。

3 求解过程

3.1 GA 的编码

GA 求解复杂问题时, 最关键的是设计一个适合问题的编码方式。只有在编码后才能进行以后的计算。一般比较直观和常规的方法是 0-1 二进制编码, 这一类编码称为常规编码, 这种编码对算法的三个算子的构造比较简单。虽然遗传算法是一种具有通用性的全局最优算法, 但如果不对问题设计算法, 其计算时间可能是非常大的。本文采用图 1 所示结构作为排课问题的遗传算法编码结构。

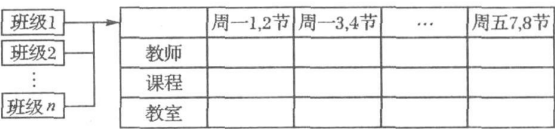


图 1 遗传算法编码结构

3.2 适应度函数的定义及计算

对于排课问题, 一般一个较好的课表应: 1) 有较高教室利用率; 2) 满足较多教师偏好; 3) 重点课程分布在较好的时段; 4) 同一门课在一周内均匀分布。对上面 4 个条件, 根据实际需要, 可以改变其对应权值的大小。同时, 为了加快算法收敛速度, 运行中每一代产生的冲突将进行不完全消除, 为了处理方便, 我们定义它为第 5 个条件, 并且, 定义一个负的权值与其对应。

单个班的适应度值函数定义为:

$$SCF_{it} = \sum_{i=1}^5 f_i \times w_i \tag{5}$$

其中 f_i 与 w_i 分别为条件 (i) 所对应的计算结果及权值。一个个体含有 n 个班级, 则该个体的平均适应度值为:

$$\overline{SCF_{it}} = \sum_{i=1}^n SCF_{it(i)} / n \tag{6}$$

其中 $FCF_{it(i)}$ 为个体中第 i 个班级的适应度值:

$$k = \sum_{i=1}^n (SCF_{it(i)} - \overline{SCF_{it}})^2 \tag{7}$$

定义个体适应度值函数为:

$$fitness = \overline{SCF_{it}} / (k^p) \tag{8}$$

参数 p 为一常量, 用于控制 k 值对适应度值的影响, 通过引入方差 k 参与适应度值的计算, 可以控制各个班级的适应度值向均值靠近, 以防造成对个别班级极不公平的现象。

3.3 初始群体的产生及冲突的消除

在初始群体产生算法中, 最常用的是随机生成初始群体的方法, 可以在一定程度上避免早熟。但是这样将会产生过多的冲突, 从而增加进化的代数。所以, 本文采用折中的办法, 产生一组没有教室冲突的个体作为初始群体。在算法运行中, 教室的数量及容量的设置同样会对算法产生较大影响, 而教室是排课系统中的一个重要资源, 在本文中, 通过教室利用率来作为考查排课系统有效性的一个重要指标, 将教室利用率乘以 10 转为适应度值参与进化计算。

$$教室利用率 = \frac{教室占用数量}{总可用教室数据} + \frac{上课学生总数}{教室总的可容纳人数} \tag{9}$$

在 GA 运行中, 由于交叉与变异操作, 均可能产生冲突, 本文采用部分消除冲突算法, 即尽量消除产生的冲突, 对于当前个体中存在的没有消除的冲突, 除了通过降低该个体的适应度

值,即加入一个负的适应度值,以降低它下次被选入的概率外,另外通过记录未消除冲突,在下次迭代过程中,作为启发式信息,有针对性地对它进行消除。而由于编码结构及其交叉与变异操作的特点,不会产生班级内课程的冲突,因此冲突主要是一个种群中个体与个体间的教师与教室冲突。本文通过将冲突个体中的一方的冲突时间段与该个体中的一空时间段进行数据交换来消除冲突,但这种交换也有可能产生新的冲突,但从概率上说,如果不存在数据异常,如一个教师课程很多的情况,产生新的冲突的概率会很低。即使产生了新的冲突,可以取消刚才的交换,重新尝试寻找另一空时间段来执行消除冲突操作,通过执行 次消除冲突操作(值依据情况进行设置,本文为 3),实验表明,可将绝大部分冲突消除。

3.4 GA的运行

遗传算法主要包括三个基本操作:选择、交叉和变异。在本文中,选择算子,利用蒙特卡罗法计算各个体的选择概率,然后采用轮盘赌选择法进行种群下一代个体的选择。对于交叉算子操作如图 2所示,通过两个个体中对应班组的互换来完成。对于变异算子,通过下面 TS的操作来完成。

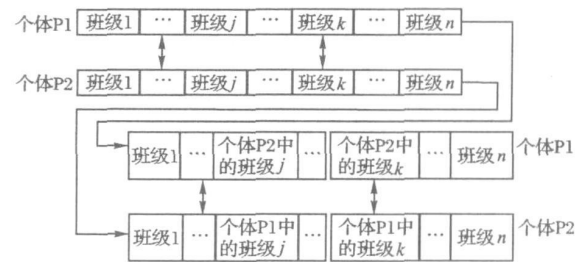


图 2 交叉算子操作

3.5 TS的操作

在本文中,TS主要用来代替变异算子,这是因为,变异算子的主要作用是使个体呈现多样性,如果对于二进制编码,往往采用一个小的变异概率,通过随机翻转某些位来实现。而对于排课问题,由于各种条件互相约束,无法实现这种随机翻转。因此,通过TS不仅可以得到比原先更优的个体解,而且也可以跳出原先的局部最优,产生新的具有更优结构的个体。

在该问题中,TS的邻域定义为相邻的两次课,为了保证每个时段的邻居数目相同,该相邻对几何意义上的相邻进行了扩展,如,周一(1,2)节课与周一(7,8)互为邻居,与周五(1,2)也互为邻居。因此,对每一时段课程均有 4个邻居。对时段按照从上午到下午、从周一到周五进行 1至 20的编号后,可以定义时段 的邻居集合如下:

$$N(i) = \{ (i+1) \bmod 20, (i+4) \bmod 20, (i-1) \bmod 20, (i-4) \bmod 20 \}$$
 (10)

上式中,mod是取模运算。

因此,如果班级 C有课程数 P则邻域解集共有 $4 \times P$ 个,取其最佳 P个解作为候选解集。禁忌表长度定义为 6,算法结束条件定义为 k次迭代运算,k值可调。与一般GA变异算子有个变异概率相同,TS的运行也需要满足一个概率。

4 实验结果及分析

本文实验数据在引用文献[3]中的实验数据的基础上,进行了改编,由于文献[3]中的课程可以单节授课,而大部分高校均采用 3节或 2节连上的方式,因此,本文采取如果有单

节情况,单节课程占有一个时段,即 2节课。因此,文献[3]中的实验数据将会导致某些班级时段不够的现象,因此本文调整了某些班级的课时数和课程数。同时,因为文献[3]不考虑教室安排的情况,因此没有教室数据,所以,本文设置了若干种不同的教室组合来对算法加以验证,表 1为引用的文献[3]数据的统计情况。其中教师数为 26

表 1 课程数据统计

班级编号	班级课程数	班级课时数	班级编号	班级课程数	班级课时数
C1	9	26	C6	9	24
C2	9	27	C7	9	28
C3	9	28	C8	9	27
C4	9	30	C9	8	27
C5	8	28	C10	9	30

表 2 WGA与 GATS适应度值

算法名称	适应度值										CPU时间
WGA	79.5	83.8	60.4	75	89.1	81.2	72.4	78.0	68.9	72.7	31.0
GATS	94.3	105.0	97.6	95.2	95.8	104.2	103.0	91.2	93.0	99.3	95

为了表述方便,将文献[3]所用算法称为 WGA算法,WGA算法通过将每个班级按 40个单元的一维数组进行编码,将所有 n个班级连成一个具有 $40 \times n$ 个单元的一维数组作为一个个体,数组内的数据为课程编号。通过计算教师课程冲突情况与课程的分布情况作为优化目标函数进行进化。

两种算法在同一台机器上多次运行,设定迭代次数为 5000,取较稳定的运行结果如表 2所示,由于 WGA不考虑教室因素,所以表 2结果,GATS已将教室所加适应度值去除,从表 2数据可以看出,本文算法的结果在适应度值以及与适应度值平均值的接近程度上都要优于 WGA算法。但由于本文的算法相比 WGA算法要复杂,考虑因素较多,因此 CPU运行时间相对较长,但对于像排课这类 NP-Hard问题,CPU时间并不是主要考虑因素。

参考文献:

[1] 王璐,邱玉辉. 基于协商的智能排课系统的研究[J]. 计算机科学, 2006 33(6):214-217

[2] 潘以锋. 高校智能排课系统的算法[J]. 上海师范大学学报:自然科学版, 2006 35(5):31-37

[3] WANG Y-Z. Using genetic algorithm methods to solve course scheduling problems[J]. Expert Systems with Applications 2003 25(1): 39-50

[4] 刑文训,谢金星. 现代优化计算方法[M]. 北京:清华大学出版社,1999

[5] 王小平,曹立明. 遗传算法——理论、应用与软件实现[M]. 西安:西安交通大学出版社,2002

[6] BARKHARD V,HAO J-K. Genetic tabu search for the multi-objective knapsack problem[J]. Tsinghua Science and Technology 2003 8(1):8-13

[7] GLOVER F,KELLY J P,LAGUNA M. Genetic algorithms and tabu search: hybrids for optimization[J]. Computers and Operations Research 1995 22(1):111-134

[8] MANTAWY A H, ABDEL-MAGID Y L, SELIN S Z. A new genetic based tabu search algorithm for unit commitment problem[J]. Electric Power Systems Research 1999 49(2): 71-78

[9] CHELOUAH R, SARRY P. Tabu search applied to global optimization[J]. European Journal of Operational Research 2000 123(2):256-270