

基于遗传算法的高校排课系统设计

陈行平 陈江 陈启华

(浙江工业大学 机电工程学院, 浙江 杭州 310014)

摘 要:分析了排课的数学模型,根据遗传算法本身特点设计了一种数据结构基因编码,及几个涉及到的优化目标评价函数,提出了利用遗传算法解决排课问题的方法,并通过试验验证了该方法的有效性、可靠性。

关键词:排课;遗传算法;多目标优化

中图分类号:TP18

文献标识码:A

文章编号:1008-293X(2004)10-0025-04

在整个教务管理系统中,学生上课、教师授课的预先安排是极其重要的一环,排课是教育计划实施的一个重要步骤.排课问题是个多目标的调度问题,涉及到众多因素如学生、课程、教师、教学资源、时间等,是运筹学中的时间表问题(Timetable Problems, 简记 TTPs).国内外在相关方面以有较为深入的研究并被公认为是 NP 难题(NP-Hard).遗传算法(Genetic Algorithm, 简记 GA)是模拟自然界生物进化过程求解极值的一类自适应人工智能技术,上世纪 60 年代由美国芝加哥大学 Holland 教授首先提出.遗传算法具有良好的并行性、很强的通用性、良好的全局优化和稳定性等优点.对于传统优化方法无法或很难解决的非线性、不可微分问题,遗传算法都能很好的解决.遗传算法的以上特点使其为解决排课问题成为可能.本文将结合实际排课中所涉及到的各类约束、优化目标设计一种特定的基因编码方法,和相应的目标评价函数,并在实际应用中取得了良好的效果.

1 数学模型分析

排课问题是求解三元组(Lecture, Time, Room)课程、时间和教室的统筹调度,即给指定的课程安排适当时间、空间等教学资源,使学校课程整体达到一个较为合理的状态.其中课程(Lecture)又包含了(Classes, Lesson, Teacher)三个元素,即每个授课的班级、课程名称和对应的老师是相关的,这些授课计划数据是系统的输入部分.而一个合理的排课结果是指使目标函数达到优值,表现为课程的时间安排均匀、教室利用率高、尽可能多满足教师偏好等.实现此结果关键在于如何解决排课过程的众多约束和有效目标函数的建立,下面就这两个问题进行具体分析.

1.1 几个典型约束

1.1.1 对于一个班级的某一个时刻不能上一门以上的课程,而在高校中由于资源紧张许多课程大班开课,因此对一次授课要求所有上该课的班级在该时刻没有其他课程.则对于任何一门课程开设时间为 t 的课程 lec 有

$$\sum_{c \in C_L} \sum_{l \in L_c} X_{clt} = 0. \quad (1)$$

其中 C_L 为参加课程 lec 所对应的所有班级集合, L_c 为班级 c 所要上的所有课程集合.

$X_{clt} = \begin{cases} 1 & \text{(班级 } c \text{ 的课程 } l \text{ 在时间 } t \text{ 开设)} \\ 0 & \text{(班级 } c \text{ 的课程 } l \text{ 不在时间 } t \text{ 开设)} \end{cases}$, 即所有班级这门课程无冲突.

1.1.2 对每个教师任何时刻只能参加一次授课,约束模型与班级约束相似,只是一次授课只有一个教师,冲突检测只需对某个教师所授所有课程集合进行统计即可,此处不在详细描述.

• 收稿日期:2004-09-21

作者简介:陈行平(1980-),男,浙江温岭人,研究方向为智能装备及其软件开发.

1.1.3 教学资源的约束,要求任何一个时刻任何一种类型的教学资源安排量不可超过学校拥有该类资源的总量;一个授课的安排要求该教学班级人数不可超过教室(或者其他教学资源)的容量.对于需要 r 类型教学资源,在 t 时刻授课的集合

$$L_{rt} = \{l \mid r_l = r, t_l = t\} \text{ 有 } |L_{rt}| \leq C_r. \quad (2)$$

其中 C_r 表示 r 类型的教学资源总量.

1.2 优化的目标

1.2.1 排课优化最为重要的目标是使得教学效果更好,即使课程安排在好的时间里上课,尤其对于相对重要的课程.本系统采取对不同的时间设定优劣系数 K_{il} (值越大的越优),对不同的课程给定优先值 M_i (值越大的为重要),则系统的这个优化目标为:

$$\text{Max}(f_1) = \sum (K_{il} \times M_i), \quad (3)$$

其中 il 为课程 i 所安排的时间.

1.2.2 在考虑教学效果的基础上,第二个重要的目标就是资源的利用率问题,一个好的安排结果可以节省大量的资源.一次授课中教学班级人数 C_{il} 与该教室容量 C_{mli} 的比值越大,则浪费越少,最大为 1,表示刚好容纳.优化目标为:

$$\text{Max}(f_2) = \sum (C_{il} \times C_{mli}). \quad (4)$$

1.2.3 其他优化目标还有课时分布均匀度、教师偏好满足程度等,根据各学校的情况可以增删,方法与前面两个相似,此处不再详细叙述.最终个体的适应度则根据各个目标值进行简单加权或者乘积所得.具体采用哪种可根据案例的情况而定,简单加权较为稳定,收敛速度较慢,乘积则相反,本系统采用的是前者,即

$$\text{Max}(f_x) = af_1 + bf_2 + \dots. \quad (5)$$

各系数通过试验确定.

2 算法设计

时间表问题的求解主要由两大模块组成,随机可行解的生成和对众多随机解进行优化,遗传算法主要作用于用后者.随机解的生成就是在可行域内随机挑选一个解,此模块还为遗传优化部分提供约束处理模块,采用面向对象技术使系统有一个良好的扩展性.

遗传算法就是借用生物进化的规律,通过繁殖、变异、竞争实现优胜劣汰,一步一步逼近问题的最优解.适合于结构优化、非线性优化、机器学习等领域.主要执行以下四个步骤:

(1) 随机建立不同可行解组成的初始群体.

(2) 计算各个个体的适应度.

(3) 根据遗传概率,利用下述操作产生新群体:①复制,利用赌盘法选出优良个体复制添加到新群体中;②交叉,利用赌盘法选出两个个体进行基因交换添加到新群体;③变异,利用赌盘法选出个体按一定概率随机变化其基因添加到新群体.

(4) 判断是否达到终止条件,达到终止输出最优解,否则转(2).

2.1 基因编码

遗传算法求解复杂问题时,最关键的是设计一个适合问题的编码方式.优秀的编码能保证系统稳定快速达到优值,劣质的编码不但影响系统性能,甚至会造成数据不完整或者优化不全面等问题,而由于求解问题本身复杂,这些由编码造成的隐患可能会隐藏在算法里.传统的二进制编码、整数编码和浮点数编码都是结构相对简单,很难表达约束众多、内部关系复杂的问题.本文提出的是一种基于结构体的编码方式.

编码结构如图 1 所示,纵向整体表示染色体,由多个不同长度的基因组成.每个横向的班级课表表示一个等位基因.交叉过程以一个等位基因为基本单位,即每次交叉整个班级的课表,以保证优秀的基因模式被遗传下来,而且大大减少由于班级内部不同课程的时间冲突;变异则针对班级课表中的某个课程对象进行.每个课程对象包括课程名称、星期几、起始节、课时数和教室等重要信息以及其他如人数、教室类型等一些附加的信息组成.为减小搜索空间,对起始节的产生是在一个模板库中随机抽取,模板库根据学校

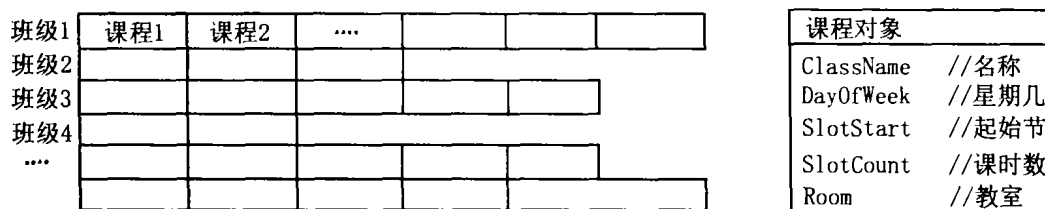


图 1 基因编码示意图

的具体情况设定.基因编码中的课表中的课程是课程对象的一个引用,也就是说不同的班级中某些课程可能是应用同一个课程对象,这使得合班课程不产生冗余信息而且检测冲突时能保证课表完整性和准确性.

2.2 选择算子

选择算子采用赌盘法原则,首先将整个群体根据个体的适应度不同分布在轮盘上,适应度大的个体占的比例多一些,适应度小的个体占的比例少一些.然后对每个个体的概率进行累积.所有个体的概率和为 100%,每个个体占其中的一个百分比段.选择时系统随机产生一个百分数,落在哪个个体的百分比段就选择哪个个体,这种选择方法对适应度高的个体选种的机会相对就多,也就是实现了优胜劣汰,但是也存在着选种适应度小的个体的可能性,这样又保证了群体的多样性,使保留在较差个体中的优秀基因段也得以保存.设计时为了得到适应度越大越容易被选中的结果,对目标函数的设计均采用求最大值的方式,如式(3),(4)和(5),而传统的优化函数一般都设计成求最小值.

2.3 交叉算子及其约束处理

交叉是遗传算法过程中一个最为重要的操作,是复杂遗传算法设计的难点.首先交叉点的选择一般有单点交叉和均匀交叉.单点交叉染色体靠近的等位基因可保持在一起,使优良的基因片断得到继承.而对于有些问题各个基因段之间相关性不是很大,则希望交叉充分,采用均匀交叉反而使收敛速度加快.考虑到排课过程各个班级之间合班授课情况较多,也即各个基因片断之间存在一定约束,采用了单点交叉的方式.

在多约束的问题中交叉算法的设计关键是如何消除约束.传统的优化的约束大多采用惩罚函数的方式,但是罚函数设计的好坏又会影响到问题是否收敛,而且增加了额外的参数.也有不少问题的求解是采用抛弃不可行解的方法,但在排课问题中稍微交叉一下就会造成冲突产生不可行解,抛弃不可行解将导致无法收敛.因此本文采用了消除冲突的策略,对交叉过程中产生冲突立即消除冲突,修正基因为有效基因,若无法消除则恢复该交叉点的现场.

3 案例分析

本系统以浙江工业大学 2004/2005(1)教学计划为算例,共有教学班级 1788 个,教师 2310 个,教室 333 个分别在两个不同校区,行政班级 974 个.种群大小为 200,迭代 5000 代,将每一代的各个目标值和整体适应度记录下来并制作成曲线图.

图 2 为最优个体适应度曲线,随着迭代次数的增加,种群的最优个体适应度平稳上升,开始的时候个体的基因随机性比较大,适应度增长迅速,到了 5000 代之后趋于平稳,系统整体达到最优的状态.图 3 为最优个体的教室利用率曲线,从中可以看出单个目标与系统整体目标不完全同步进行,教室利用率目标到

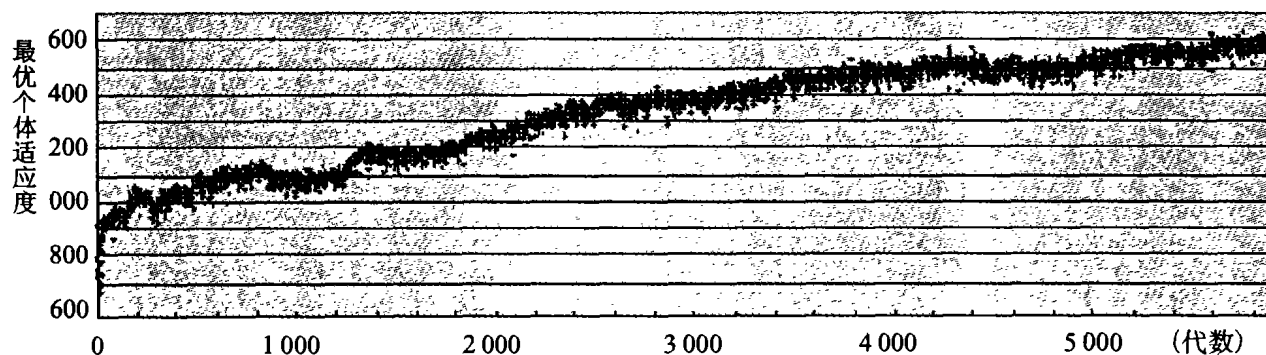


图 2 种群最优个体适应度曲线图

3 000 代左右已经趋于稳定,但由于其目标函数的作用,系统继续迭代该目标仍然保持水平.从图中可以看出,刚开始迭代时稍有振荡,这是由于随机生成的个体各个目标有一定的倾向性,不完全随机.趋势变化从试验角度验证了该算法的可行性,说明了遗传算法在时间表问题求解是可行的方法.

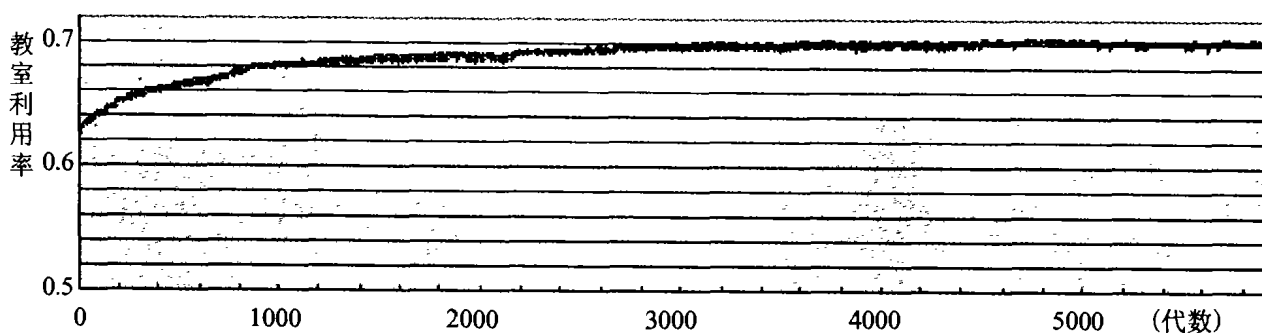


图 3 最优个体教室利用率曲线图

参考文献:

- 1 何永太.二部图在排课系统设计中的应用[J].安徽水利水电职业技术学院学报,2003,2(3):45-47.
- 2 郑伟华,郑金华.狭义遗传算法的遗传机理分析[J].湘潭大学自然科学学报,2003,25(1):21-23.
- 3 杨宇.高校排课系统理论与开发——遗传算法在课表问题中的应用[D].北京:北京理工大学,2003.
- 4 张春梅,行飞.用自适应的遗传算法求解大学课表问题[J].内蒙古大学学报(自然科学版),2002,33(4):459-464.

Application of Genetic Algorithm in Timetabling Problem Chen Hangpin

Chen Jiang Chen Qihua

(College of Mechanical and Electric Engineering, Zhejiang University of Technology, Hangzhou, Zhejiang, 310014)

Abstract: This paper analyses the mathematic model of timetabling problem, and designs a data structure of gene and some evaluating functions of the goals. A solution to timetabling problem via genetic algorithm is provided. The validity and reliability are proved by experiments.

Key words: timetabling problem; genetic algorithm; multi-objective optimization