

基于 PBIL 算法的分层教学自动组班研究^①

刘日仙¹, 袁利永²

¹(金华职业技术学院, 金华 321017)

²(浙江师范大学 数理与信息工程学院, 金华 321004)

摘要: 分层教学模式的采用对选课排课工作带来了新的挑战。提出了一种基于 PBIL 的分层教学自动组班算法, 重点论述了基因结构的设计, 以及目标优化函数与自动组班约束条件之间的关系。基于实际数据的实验测试表明, 本文提出的自动组班算法能够较好地解决分层教学模式下产生的排课选课问题, 实际应用效果良好。

关键词: PBIL; 组合优化; 进化计算; 基因算法; 自动组班

Automatic Grouping of Hierarchical Teaching Based on PBIL Algorithm

LIU Ri-Xian¹, YUAN Li-Yong²

¹(Jinhua Polytechnic College, Jinhua 321004, China)

²(College of Mathematics Physics And Information Engineering, Zhejiang Normal University, Jinhua 321004, China)

Abstract: The adoption of hierarchial teaching mode brings a new challenge to course scheduling and course elective. In this paper, an algorithm for automatic grouping of hierarchical teaching based on PBIL is proposed. The design of gene structure and the relationships between the optimal function and automatic grouping class constraint is emphasized. Experimental tests on real data show that the proposed algorithm of automatic grouping class can successfully solve the course scheduling problem in the hierarchical teaching mode. The practical application of it has good results.

Key words: PBIL; combinatorial optimization; evolutionary computation; genetic algorithm; automatic grouping class

分层教学是目前许多高校正在推进的一项教学改革, 尤其是针对一些全校性的公共课程。然而, 这项改革也为教务管理工作带来了重大的挑战, 其中最棘手的是排课选课问题。例如我校目前正在进行的公共计算机分层教学改革: 新生入校后首先进行分层考试, 然后根据分层情况开设若干个教学班级供学生进行选课, 但由于其它课程在此之前已经选定, 学生要在空闲时间段内找到合适的计算机教学班级并不容易(选课成功率仅 60%左右)。为了解决这一问题, 我们提出由学生上报空闲时间段, 然后通过相关算法优化组合出相应的教学班级, 使得尽可能多的学生能够顺利修读计算机课程。

基于群体的增量学习算法(Population Based Incremental Learning, PBIL)是进化学习算法的一种, 首先由美国卡耐基梅隆大学的 Baluja 在 1994 年提出^[1],

随后有不少论文对其理论和应用进行了研究。文献[2]提出了 PBIL 算法的马尔科夫模型, 并证明 PBIL 算法的收敛性。文献[3]将 PBIL 中的基因位由原来的二进制形式扩展成十进制形式, 并提出用信息熵作为结束条件的判据。为了提高 PBIL 算法的收敛能力, 文献[4]提出一种采用多概率向量的 PBIL 算法。由于 PBIL 算法具有运行过程简单、解决问题快速准确、对复杂约束条件的增减具有较好的适应性等优点, 因而引起广泛关注, 并被成功运用于解决各种不同领域的优化问题^[5-8]。因此, 本文提出采用 PBIL 算法求解分层教学模式下的自动组班问题。

1 PBIL 算法简述

PBIL 算法将原优化问题的解空间 S 映射到基因学习算法能够操作的基因解空间 S^1 。可行解 $s^1 \in S^1$ 由 1

① 基金项目: 2009 年度浙江省教育厅项目(Y200908315); 2010 年度浙江省教育厅项目(Y201016495)

收稿时间: 2011-04-14; 收到修改稿时间 2011-05-06

个基因位组成,其中第*i*位的值可以是集合 $\{a_{i1}, a_{i2}, \dots, a_{ini}\}$ 中的任意一个或多个。集合 $\{a_{i1}, a_{i2}, \dots, a_{ini}\}$ 称为第*i*位的等位基因。如果第*i*位只能取 $\{a_{i1}, a_{i2}, \dots, a_{ini}\}$ 中的一个值,则称第*i*位为单等位基因位。对于有1个基因位、每个基因位都可以从*n*个等位基因中取值的优化问题,其基因解可以用一个布尔矩阵 $S_{bl} \times n$ 来表示,如果矩阵元素 $s_{ij}=1$,就表示第*i*个基因位的取值为该基因位的第*j*个等位基因值。

向量 $P^l = (P_1, P_2, \dots, P_l)$ 用来表示在解空间 S^l 的1个基因位中每个基因位取与该基因位对应的等位基因值的概率。其中 $P_i = (p_{i1}, p_{i2}, \dots, p_{ini})$ 表示第*i*个基因位取与该基因位对应的*n_i*个等位基因值的概率。 p_{ij} 为第*i*个基因位取等位基因集合 $\{a_{i1}, a_{i2}, \dots, a_{ini}\}$ 中第*j*个值的概率, $j = 1, 2, \dots, n_i$,且每个基因位取等位基因的概率之和都为1。

PBIL算法把优化选择问题描述为 $\min f(s), s \in$ 可行解空间, f 为优化目标函数。

典型的PBIL进化学习过程描述如下:

步骤1: 初始化学习概率 $P: P_{ij} = 1/n_i, 1 \leq j \leq n_i, P_{ij} \in P_i, 1 \leq i \leq l$;

步骤2: 由学习概率 P 指导产生*n*个解 s_1, s_2, \dots, s_n ;

步骤3: 计算*n*个解的目标函数值 $f(s_1), f(s_2), \dots, f(s_n)$,并找出最优解 s' ;

步骤4: 用最优解向量 s' 修正学习概率 $P: p_{ij} = p_{ij} + \psi \times s'_{ij}, p_{ij} = p_{ij} / \sum_{j=1}^{n_i} p_{ij}, 1 \leq i \leq l$;

步骤5: 返回步骤2,直至满足一定的结束条件为止。

参数 ψ 是一个经验常数,在不同规模和约束条件下,它的取值对于问题求解速度和成功率都有较大的影响,一般介于0.01~0.1。

为了防止学习概率过早地收敛到0或1附近而产生早熟现象,Baluja.S还将遗传算法中的变异操作引入到PBIL算法中^[9],在每一代的最优解对学习概率进行修正后,再按变异率 P_m 随机地选择部分学习概率值 P_i 进行随机扰动, R_m 是变异速率。扰动方法可表示为: $\text{If}(\text{Random}(0,1) < P_m) P_i = P_i * (1.0 - R_m) + \text{Random}(0 \text{ Or } 1) * R_m$ 。

2 基于PBIL的自动组班算法研究

2.1 自动组班问题描述

在分层教学模式下,自动组班问题主要涉及以下

三个对象:学生、教室、教师。自动组班算法需要解决的主要问题是:在满足教室、教师时间约束的前提下,求解教学班的上课时间,使得尽可能多的学生能够修读当前课程。

自动组班的冲突情况一般包括:

(1) 指定时间段的教学班数量不能超过可用教室资源数量。

(2) 指定时间段的教学班数量不能超出可用教师资源数量。

以上2条约束条件称为硬约束条件,这是在自动组班时必须满足的条件。除此之外,算法还应满足以下几方面的需求:

(1) 让尽可能多的学生能够修读当前课程。

(2) 每个教学班多个上课时间段之间需要有合理的时间间隔。

(3) 各教学班级规模(学生人数)尽可能平衡。

以上的约束条件成为软约束条件,这是在自动组班时需要尽量满足的条件。

自动组班问题所涉及的资源描述如下:

设 $S\text{Set} = \{S_i | 1 \leq i \leq L\}$ 为学生集, $ST\text{Set} = \{ST_i | ST_i$ 表示学生 S_i 的时间占用情况, $1 \leq i \leq L\}$ 为学生时间集,其中 $ST_i = \{ST_{i1}, ST_{i2}, \dots, ST_{iz}\}, ST_{ij} \in \{0, 1\}$, $ST_{ij} = 1$ 表示该时间段有空, $ST_{ij} = 0$ 表示该时间段占用, L 为学生数量, Z 为一周时段数。设 $C\text{Set} = \{C_i | 1 \leq i \leq M\}$ 为教室集, $CT\text{Set} = \{CT_i | CT_i$ 表示教室 C_i 的时间占用情况, $1 \leq i \leq M\}$ 为教室时间集,其中 $CT_i = \{CT_{i1}, CT_{i2}, \dots, CT_{iz}\}, CT_{ij} \in \{0, 1\}$, M 为教室数量。设 $T\text{Set} = \{T_i | 1 \leq i \leq N\}$ 为任课教师集, $TT\text{Set} = \{TT_i | TT_i$ 表示教室 T_i 的时间占用情况, $1 \leq i \leq N\}$ 为教师时间集,其中 $TT_i = \{TT_{i1}, TT_{i2}, \dots, TT_{iz}\}, TT_{ij} \in \{0, 1\}$, N 为教室数量。设 K 为教学班一周所要占用的时段数量(本文实验课程一周上2次课,故 $K=2$),设 P 为需要开设的教学班级数量。

2.2 算法设计

自动组班问题是多资源组合优化问题,它是属于一类NP难问题。在分层教学模式下,教学班上课时间安排是算法最复杂的部分,其次是教师和教室的安排。为了降低算法复杂度,提高算法效率,可将自动组班算法分解为以下几个步骤:第一步,在满足教室和教师总体约束条件下,求得满足学生时间约束的满意解 ZB^{ST} ;第二步,基于 ZB^{ST} 求解满足教师时间约束的最优解 ZB^{TT} ;第三步,基于 ZB^{ST} 求解满足教室时间约束的最优解 $ZBCT$,最后将

ZBST、ZB^{TT}、ZB^{CT} 合并为 ZB 即可。由于 ZBST 已经满足教室及教师总体约束要求，为教学班分配教室和教师则变得较为简单，因此，本文只重点介绍求解 ZBST 的算法设计。

设 TC={TC₁,TC₂,TC₃,...,TC_P} 为拟开设的教学班集合，自动组班问题的解可描述为由 K*P 个基因位构成的基因序列 ZBST={ZBST₁,ZBST₂,ZBST₃,...,ZBST_{K*P}}，其中 ZBST_i 表示第((i-1)\K + 1)个教学班第((i-1) Mod K + 1)个时段的上课时间。为了简化计算，将一周的 Z 个时间段集合{WT}按顺序划分为互不重叠的 K 份，即 {WT}={WT₁} ∪ {WT₂} ∪ .. {WT_K}，其中某个教学班第 i 个时段的时间取值于{WT_i}。概率矩阵 PST(K*P)×Z 表示某个教学班时段安排在某一时间的概率，如 PST_{ij} 表示第((i-1)\K + 1)个教学班第((i-1) Mod K + 1)个时段采用上课时间 j 的概率。

求解 ZBST 的目标函数表示为 f^{ZB}=w₁*(f^{CT}+f^{TT})+w₂*fST+w₃*f^{FB}，其中 f^{CT} 对应于教室总体约束的罚函数，f^{TT} 对应于教师总体约束的罚函数，fST 对应于学生组班失败的罚函数，f^{FB} 对应于教学班多时段上课时间间隔约束的罚函数。w₁ 分别对应两条硬约束，而 w₂,w₃ 对应软约束，因此 w₁>>w₂,w₃。w₂,w₃ 的值则根据约束条件的倚重程度进行设置。

$$f^{CT} = \sum_{i=1}^Z (NeedC(i) > HaveC(i) ? NeedC(i) - HaveC(i) : 0)$$

其中 NeedC(i) = $\sum_{j=1}^{K*P} (ZB_j = i ? 1 : 0)$ 表示第 i 个时段需要的教室数量，HaveC(i) = $\sum_{j=1}^{K*P} CT_{ij}$ 表示第 i 个时段空闲教室数量。

$$f^{TT} = \sum_{i=1}^Z (NeedT(i) > HaveT(i) ? NeedT(i) - HaveT(i) : 0)$$

其中 NeedT(i) = $\sum_{j=1}^{K*P} (ZB_j = i ? 1 : 0)$ 表示第 i 个时段需要的教师数量，HaveC(i) = $\sum_{j=1}^{K*P} TT_{ij}$ 表示第 i 个时段空闲教师数量。

$$f^{ST} = \sum_{k=1}^L (IsMatch(S_i) ? 0 : 1)$$

其中 IsMatch 函数含义如下：若存在 TC_x ∈ TCSet, 对于任何 j ∈ {1..K}, 令 t=ZB(x-1)*K+j, 都有 STit=1, 则返回真；否则返回假。

$$f^{FB} = \sum_{i=1}^P \sum_{j=1}^{K-1} (ZB_{j+1} - ZB_j < \varepsilon ? \varepsilon - ZB_{j+1} + ZB_j : 0)$$

ε 为教学班各上课时段时间间隔控制参数。

为使各教学班人数尽可能均衡，采用的方法如下：

设 AC(s)={C_i|学生 s 可组班到教学班 C_i 中}，将学生 s 组班到第 h* 个教学班，使得 h* 满足下列条件：

$$h^* = \arg_h \min \{count(C_h) | C_h \in AC(x)\}$$

2.3 算法实验与效果分析

为了验证本文算法的有效性，采用 VB 语言对算法进行了实现，采用本校 2010-2011 学年第一学期修读《计算机应用（理）》课程的学生作为测试数据，其中待选课学生 1978 人，平均空闲时间为 25.4 节。根据经验，PBIL 算法中的种群大小取 50，概率修正参数 ψ 取 0.02。测试环境采用 Windows XP 操作系统、3.2G P4 CUP、2G 内存。在测试中对一些次要约束条件(如教室类型、任课教师特殊要求等)进行了简化，测试结果如表 1 所示。

表 1 算法实验测试结果

教学班数量	组班成功率 (平均)	班级均衡率 (平均)	运算时间 (平均)
10	97.92%	99% 以上	156 秒
15	99.79%	99% 以上	129 秒
20	99.92%	99% 以上	98 秒
25	100.0%	99% 以上	85 秒
30	100.0%	99% 以上	78 秒

3 结语

本文分析了分层教学模式给选课排课工作带来的影响，提出采用 PBIL 算法求解分层教学模式下的自动组班问题。基于实际数据的实验测试表明，本文提出的自动组班算法能够较好地解决分层教学模式下产生的排课选课问题，实际应用效果良好。

参考文献

- 1 Baluja S. Population-Based incremental learning. Technical Report, CMU-CS-94-163, Carnegie Mellon University,1994.
- 2 Rastegar R, Hariri A, Mazoochi M. A Convergence Proof for the Population Based Incremental Learning Algorithm. Proc. of the 17th IEEE International Conference on Tools with Artificial Intelligence. Hongkong, China, 2005:387-391.
- 3 金炳尧, 蔚承建, 何振亚. 一个用于优化搜索的学习算法. 软件学报, 2001, 12(3):448-453.
- 4 Zhang QB, Wu TH, Liu B. A Population Based Incremental Learning Algorithm with Elitist Strategy. Proc. of the 3rd

(下转第 220 页)

```

ipv6 address 2001:DA8:FF3A:C8CA:100::/64
ipv6 nd ra prefix 2001:DA8:FF3A:C8CA::/64 3600
3600 no-autoconfig
undo ipv6 nd ra halt
ipv6 nd autoconfig managed-address-flag
ipv6 nd autoconfig other-flag
    
```

配置中可见, 由于 DHCPv6 服务器无法向 IVI 主机分配 IPv6 网关, 因此以 RA 的方式向 IVI 主机指定缺省网关, 同时禁用了 IPv6 地址的无状态自动分配功能, 确保各 IVI 主机能够得到合法的 IVI 地址。

4 IPv6实验设计与实例

我们建立了一个基于纯 IPv6 的 WWW 网站, 作为 IPv6 实验教学和有关科研信息的交流。在该网站上定期发布 IPv6 有关知识、最新 IPv6 科研动态、相关实验教学设计, Web 服务采用了 Windows 平台下的 IIS 实现。

根据教学与科研的实际需要, 我们设计了相关的 IPv6 实验, 包括 IPv6 协议支持、IPv6 应用、IPv6 路由协议和 IPv6 高级实验四个类别^[4], 每个类别的具体内容如表 1 所示。这些实验已应用于我们开设的“网络与系统原理”、“计算机网络”、“IPv6 与下一代互联网”等多门计算机网络相关课程的实验教学内容。此外, 我们还在该实验环境下开展了基于 IPv6 协议测试, 性能测试等科研项目。

3 结语

通过构建科学、合理的纯 IPv6 实验环境, 可以降低 IPv6 实验的复杂性, 简化实验设计, 提高实验的效率。我们构建的纯 IPv6 实验环境中约有 200 台纯 IPv6 主机, 所有主机通过纯 IPv6 链路连接到 CERNET2 华南核心节点,

(上接第 227 页)

International Conference on Natural Computation. Haikou, China, 2007,2:583-587.

5 Gallagher M, Freaun M. Population Based Continuous Optimization Probabilistic Modeling and Mean Shift. Evolutionary Computation, 2005,13(1):29-42.

6 袁利永,倪应华,金炳尧,马永进.PBIL 算法在组合优化问题中的应用研究.计算机工程与科学,2011,33(3).

同时以 IVI 技术实现与 IPv4 主机的互通, 并通过 IPv6 Web 网站进行信息的发布和共享。该实验环境向学生和教师提供了一个进行纯 IPv6 学习、实验和研究的环境, 满足下一代互联网相关课程教学、实验和科研的需要。实践证明, 学生在该环境下可以完成各种基于 IPv6 的网络实验, 帮助他们将理论与实践相结合, 更好地提升其能力; 老师在该环境下可以进行各种 IPv6 网络科研和学习, 帮助他们更好地从事下一代互联网的网络技术研究和应用系统开发。

表 1 IPv6 实验内容表

实验类别	实验内容
IPv6 协议支持	IPv6 主机建立实验, IPv6 主机通信实验, IPv6/v4 互通实验、各操作系统 IPv6 协议支持命令
IPv6 应用架设	DNS 解析、WWW、FTP IPv6 的支持
IPv6 路由协议	IPv6 静态路由、RIPng、OSPFv3 和 BGP4+配置、策略路由
IPv6 高级实验	IPv6 协议分析, IPv6 组播、IPv6 协议测试, IPv6 防火墙, IPv6 网络性能

参考文献

1 江魁,张凡.园区网中的 IPv6 实现及应用.计算机系统应用,2005,14(5):35-38.

2 李星.IPv6 现状、研究动态和过渡策略.CERNET NOC 会议, 2010.

3 郭晓东,郭汝廷.实践 IVI-山东大学 IPv6 网络部署方案.中国教育网络,2009,4:12-14.

4 江魁,龚巧华,杨文玲,张园红.基于 IPv6 的网络课程实验教学初探.计算机科学,2007,34(7):49-51.

7 袁利永,金炳尧,刘日仙.基于 PBIL 算法的高校自动排考系统研究.计算机系统应用,2010,19(5):205-208.

8 袁利永,金炳尧,曹振新.PBIL 算法求解物流中心选址优化问题研究.计算机系统应用,2010,19(11):242-245.

9 Baluja S, Caruana R. Removing the genetics form the standardgenetic algorithm. Proc. of the International Conference on Machine Learning. San Mateo, 1995, 38-46.