

基于混合遗传算法的任务驱动分组优化研究

李浩君 杜兆宏 邱飞岳

(浙江工业大学教育科学与技术学院 杭州 310023)

摘要 智能算法应用到教学领域来实现自动分组具有重要意义。针对网络学习环境下任务驱动教学中如何按最优分组方案进行小组划分的问题,综合考虑了分组问题中学习者之间的特征差异和任务难易程度等影响因素,构建了基于任务驱动分组优化问题的数学模型,提出了基于混合遗传算法的任务驱动分组优化策略。在 MATLAB7.0 平台上,运用混合遗传算法对任务驱动的分组优化进行了仿真实验。实验结果表明,基于混合遗传算法的任务驱动分组优化是可行且有效的。

关键词 协作学习分组,任务驱动,分组优化模型,混合遗传算法

中图分类号 TP391.41 文献标识码 A

Optimized Research for Task-driven Grouping Based on Hybrid Genetic Algorithm

LI Hao-jun DU Zhao-hong QIU Fei-yue

(College of Education Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China)

Abstract Intelligent algorithm that applies to the education field to realize automatic grouping has great significance. In the task-driven teaching under the network learning environment for how to group divided according to the optimal grouping scheme, the factors of the characteristic differences between learners and the degree of task difficulty were considered, a mathematical model based on task-driven grouping optimization problem was built, the strategy of task-driven grouping optimization based on hybrid genetic algorithm was proposed. We had done an simulation experiment by using hybrid genetic algorithm on MATLAB7.0 platform. Experimental results show that the optimization of task-driving grouping based on hybrid genetic algorithm is feasible and effective.

Keywords Collaborative learning group, Task-driven, Grouping optimization model, Hybrid genetic algorithm

1 引言

随着信息技术的快速发展,学习者之间通过网络协作学习来完成学习任务已成为一种新型的学习方式。在网络协作学习中,根据学习者的特征进行合理有效的分组对于高效地完成学习任务显得十分重要。

通过对国内外有关协作学习分组文献的研究发现,随机分组由于未考虑学习者之间的个性特征差异和学习任务难易程度等因素,分组学习效率较低;穷举法虽然考虑到了学习者之间的个性特征差异,但其分组效率同样较低^[1]。刘菊香提出通过模糊聚类方法将测量出的学习粉盒数值作为样本进行聚类,从中找出相似学习风格的学习者进行分组^[2]。Sung CT 等采用基于距离的 RMH(random mutation hill climbing)算法寻找一群给定学生的最优分组^[3]。马利东提出使用模糊聚类方法实现“寻找最佳的学习伙伴”和“最佳的分组”的算法模块^[4]。陈瑶等提出了两点组群杂交以及基于适应值的组群启发式变异遗传算子,实现了面向组群的进化过程^[5]。Lin 等提出了基于改进的粒子群算法来解决协作学习分组问题^[6]。Hwang 等提出了基于遗传算法的分组模型^[7]。胡慧等提出了基于改进蚁群算法的协作学习分组模型^[8]。

Moreno 等提出了一种利用改进遗传算法来解决分组问题的策略^[9]。王中华提出了在网络环境下基于聚类分析的分组策略,其在每类中按比例抽调成员来组成协作学习小组^[10]。近年来,针对网络协作学习自动分组的研究较多,但针对任务驱动的学习分组的研究却很少。

本文结合典型的组合优化思想,在综合考虑任务驱动分组中学习者之间个性特征差异和任务难易程度等影响因素的基础上,构建了网络学习环境下任务驱动分组优化的数学模型;并且在 MATLAB7.0 仿真平台上利用混合遗传算法和传统遗传算法对任务驱动分组优化的可行性和有效性进行了仿真对比验证。

2 任务驱动的分组问题分析

任务驱动是一种建立在建构主义教学理论基础上的教学方法,具体而言就是以具体学习任务进行诱导,不断加强和维持学习者的学习动机,通过完成学习任务的方式掌握学习的核心内容,并以展示完成任务成果的方式来体现教学的效果^[11]。如果在网络协作学习环境中能根据学习者的个性特征和学习任务的难易程度进行有效地分组,则可以创设更好的学习氛围,激发学习者积极地进行讨论并提出问题,从而提

本文受 2016 年国家社科基金年度项目:移动设备知识传播的情景感知服务机制及运行实证研究(16BTQ084)资助。

李浩君(1977-),男,博士生,副教授,主要研究方向为智能计算、移动学习, E-mail: zgdhjl@zjut.edu.cn;杜兆宏(1987-),男,硕士生,主要研究方向为智能计算、个性化学习;邱飞岳(1965-),男,博士,教授,主要研究方向为教育信息化、智能算法。

升学习效果^[12]。任务驱动学习中的任务正好为协作学习提供了前提条件。

在网络学习环境下的任务驱动分组协作学习中,如果学习者的个性特征正好能很好地与专业教师(或学科专家)所设定的任务要求相匹配,则不但可以促使学习者更加高效地完成其所分配的学习任务,而且还能促使学习者积极地参与到协作学习任务中并自如地施展个人的才华,同时还能在完成分组协作学习任务的过程中不断锻炼协作能力,培养协作学习习惯,建构个人的知识体系结构等。

可将该领域中的任务驱动分组优化问题视为组合优化问题来实现。本文重点关注的是不同难易程度的学习任务和不同个性特征的学习者之间形成最佳的匹配,即针对某个需要学习者之间相互协作才能完成的学习任务,根据分组原则和学习任务难易程度等,从待分组的 learners 中找出与特定任务最佳的匹配,以得到最优的分组方案。

3 任务驱动分组优化模型的构建

文献[13]中提出了组合优化问题中解决多维背包问题进行数学模型;文献[7]中提出了基于遗传算法的分组模型;文献[8]中提出了基于改进蚁群算法的协作学习分组模型。以上这些案例为本文构建任务驱动的分组优化的数学模型提供了一定的借鉴。

通过对典型组合优化问题之一的多维背包问题进行分析研究发现,其核心思想是要把 n 个价值为 $v_j (j=1, 2, \dots, n)$ 的物品选择装入 m 个容积为 $c_i (i=1, 2, \dots, m)$ 的容器中,使装入容器的物品总价值最大,即目标函数求得最优解。就其本质而言,有两个约束条件,即容量约束和每个物品只能装入一个容器的 0/1 规划问题,这在计算理论上属于完全问题,计算的复杂性为 $O(2^n)$ 。分组问题同样也是在满足一定的约束条件下,将一个包含多元素的集合划分为几个互不相交的子集组合在一起,以使目标函数得到最优解的一类问题^[9]。以上所述的这两个问题在本质上是一样的。

本文将网络协作学习中的任务驱动分组优化问题抽象为多维背包问题来实现。在实施任务驱动的分组教学过程中,要将待定分组的 n 名不同个性特征的学习者分配到对应的 m 项难易程度不同的学习任务中,以形成任务驱动的协作学习小组的划分,进而以小组协作的形式来完成专业教师或学科专家设定的学习任务,其目标就是在组合优化问题中的可行解中求出最优解,即求出最优的分组方案。

有研究表明,当协作学习活动要求完成活动的任务比较简单时,采用同质分组原则效率较高;当要求完成的任务比较复杂时,采用异质分组原则活动效率较高^[11]。考虑到实际网络学习环境中学习者学习能力水平的差异较大、学习任务的难易程度各异等因素,本文采用异质分组和同质分组相结合的原则对学习者进行分组。

在网络学习环境中,任务驱动分组优化的数学模型描述如下:

$$\lambda_k = \begin{cases} \sum_{i=1}^{n-1} \sum_{j=i+1}^n d_{ij}(x_{ik}w_i)(x_{jk}w_j), & g=1 \\ -\sum_{i=1}^{n-1} \sum_{j=i+1}^n d_{ij}(x_{ik}w_i)(x_{jk}w_j), & g=0 \end{cases} \quad (1)$$

其中, $k=1, 2, \dots, m, m$ 表示所设定的学习任务总数; $w_i, w_j \in$

$w[n^2]$ 是满足特定学习任务需求的学习者 i 和 j 在当前任务中的优先级, $w_k, w_j \in [0, 1]$, 这里以 0/1 决策变量来判定学习者 i 和 j 是否被特定的学习任务 k 所选择, 如果被选择, 则以 1 表示, 否则以 0 来表示; $w[n^2]$ 为所有待分组学习者的权值求交集和并集后得到的汇总向量。

分组原则在式(1)中用变量 g 来判断, $g=1$ 表示采用异质分组原则, $g=0$ 表示采用同质分组原则。其中按同质分组原则进行求解, 所求得的结果为差异的最小值, 而模型中目标函数是求最大化问题, 因此式(1)中当 $g=0$ 时, 需乘以 (-1) 将其变换成求最大化问题。

$$d_{ij} = \sum_{k=1}^c |v_k[i][Factor] - v_k[j][Factor]| v_k[Weight] \quad (2)$$

式(2)用来计算各加权学习任务 k 中学习者之间的特征差异。其中, $i \in (1, 2, \dots, c), c$ 表示第 k 项任务所选取学习者的特征数; d_{ij} 表示学习者 i 和 j 之间各项特征差异所求得的总和。

$$\text{Max} \sum_{k=1}^m P_k \lambda_k \quad (3)$$

式(3)为目标函数。其中, $k=1, 2, \dots, m, m$ 表示所设学习任务的总数目; P_k 则表示为给当前第 k 项学习任务所赋予的优先级。模型中目标函数的含义为: 共有 m 项以优先级 P 加权的任务, 其中第 k 项协作学习任务需要的组员数为 s_k , 现共有 n 个学习者可供分配; 以 w_i 表示学习者 i 在当前学习任务中的加权值, 以 w_j 表示学习者 j 在当前学习任务中的加权值, 以 d_{ij} 表示学习者 i 和 j 之间各项特征差异所求得的总和, 依据特定学习任务 k 所选择的分组原则求 λ_k , 使得优先级为 P_k 的学习任务的 λ 之和为最大值, 即得到最优的分组方案。

$$\text{s. t.} \begin{cases} \sum_{i=1}^n x_{ij} = s_j, x_{ij} \in \{0, 1\}, j=1, 2, \dots, m & (4) \\ \sum_{j=1}^m x_{ij} = 1, x_{ij} \in \{0, 1\}, i=1, 2, \dots, n & (5) \end{cases}$$

式(4)和式(5)为约束函数。约束条件式(4)为任务组容的限制, 其中第 i 个加权任务的组容表示为 s_i 。在任务驱动的分组问题中, 如果将待定分组的学习者分配到了某个学习任务中, 那么每位学习者就要占据组容一个名额。约束条件式(5)表示每个待分配的学习者只能分给所设定学习任务总数中的一项。

4 基于混合遗传算法的任务驱动分组优化策略

智能算法为求解优化问题提供了一种全新的手段。遗传算法^[14](Genetic Algorithms, GA)是一种结合自然界的遗传机制和生物进化论的并行随机搜索最优化方法。由于组合优化问题具一些独特的结构特性, 因此遗传算法(GA)比较适合用来求解组合优化问题, 并且由于其在求解组合优化问题中表现出了良好的全局搜索性能而被很多研究者所认同。

但利用传统遗传算法(GA)求解具有一定约束条件的组合优化问题时, 在采用二进制编码的方式下, 其在对具体问题优化的过程中会出现不可行解的情况。针对这一问题, 文献[13]提出使用惩罚法强制将不满足约束条件的不可行解转化为可行解。但在某些求解具体问题的过程中, 使用这种方法会出现种群个体的适应度为 0 的可能, 从而导致迭代效果不理想; 另外, 遗传算法(GA)作为一种全局搜索最优化法,

在全局搜索方面的能力很强,但在局部搜索方面表现较差。针对传统遗传算法(GA)上述的不足,本文选择采用贪婪算法(Greedy Heuristics,GH)和传统遗传算法(GA)相结合的混合遗传算法(Hybrid Genetic Algorithm,HGA)来对任务驱动的分组优化问题进行求解。在求解任务驱动的分组优化问题的过程中,利用贪婪算法将种群的初始化和遗传操作中产生的不满足约束条件的不可行解修复成可行解^[15]。通过传统遗传算法的寻优,在进化过程中不断执行选择、交叉、变异以及贪婪算法对群体进行修正等一系列过程,最终所求问题将不断优化,从而得到最优解。

其具体进化过程如图 1 所示。

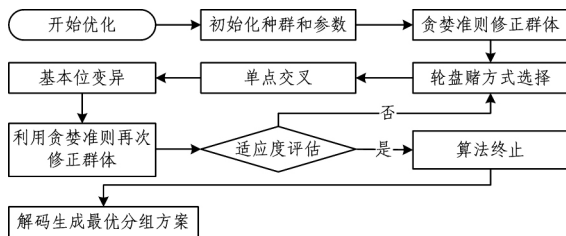


图 1 基于混合遗传算法的任务驱动分组优化的流程图

利用混合遗传算法对具体问题求解时的主要操作步骤如下。

(1) 编码

在任务驱动的分组优化问题中,将 n 个个性特征值为 v_i 的学习者按顺序排列,以此构成任务驱动的分组问题的遗传编码。本文以二进制编码的方式来表示基因,编码中的 1 表示将符合此项学习任务要求的学习者分到该任务中,0 则表示不把该学习者分到该学习任务中,以此来区分学习者是否被分到对应的学习任务中。

(2) 适应度函数的确定

本文将模型中的目标函数直接作为适应度函数来对群体的适应度进行评估,在进化过程中评估 m 项以 P 加权的学习任务。其中第 k 项学习任务需要的组员总数为 s_k ,学习者 i 在该项学习任务中的加权值为 w_i ,学习者之间特征差异的总和为 d_{ij} ,从 n 个待分配的学习者中依据特定学习任务 k 的分组原则求 λ_k ,使优先级为 P_k 的学习任务的 λ 之和为最大值,即目标函数求得最优解。

(3) 选择算子

选择又称复制,是在群体中选择生命力强的个体产生新的群体的过程。本文采用轮盘赌方式进行选择操作,即每一个个体进入下一代的概率等于它的适应度值与整个种群中个体适应度值的比例。适应度值越大,其被选中的可能性就越大,进入下一代的概率就越大。

(4) 交叉算子

交叉又称重组,是以较大的概率从群体中选择两个个体,交换两个个体的某些位。本文采用单点交叉,即随机选取一点作为基因的交叉点。

(5) 变异算子

变异是将个体染色体编码串中的某些基因座上的基因值用该基因座的其他等位基因来替代,从而形成一个新的个体。本文采用基本位变异,即对个体编码串中以变异概率随机制定的某一位或某几位基因座上的值做变异运算。

5 实验仿真与验证

为了验证基于混合遗传算法的任务驱动分组优化策略的可行性和有效性,在 MATLAB7.0 仿真平台上分别用混合遗传算法(Hybrid Genetic Algorithm,HGA)和遗传算法(Genetic Algorithm,GA)对这一问题进行仿真对比验证。

在选取实验对象时,由于本文针对的是网络协作学习环境下的学习者分组,而且采用的是异质分组和同质分组相结合的分组原则,因此选取的实验对象为网络协作学习环境下某门课程需要协作完成学习任务的网络课程完成注册的不同个性特征的学习者。在本实验中,网络协作学习课程中作为驱动的学习任务是由专业教师或学科专家事先根据网络学习者的实际情况和特点所设定的,这样的设计使得网络协作学习中要完成的协作学习任务具有了层次性,从而能更好地与不同个性特征的网络学习者形成最佳匹配。

为了分析不同参数对分组优化效果的影响,分别设定了两组不同的参数进行实验仿真。表 1 为设定的两组参数值列表。

表 1 算法实验参数

Gen	Pop	P_c	P_m
100	20	0.50	0.01
150	30	0.60	0.02

在 MATLAB 中编写程序来对任务驱动的分组优化过程进行仿真。首先选取进化代数为 100、种群为 20、交叉概率 $P_c=0.5$ 、变异概率 $P_m=0.01$ 来进行优化仿真实验。HGA 算法(三角)和 GA 算法(方块)的优化曲线如图 2 所示。

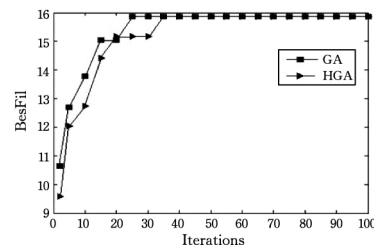


图 2 进化曲线对比图(一)

如图 2 所示,随着进化代数的增加,最优解的值整体呈上升趋势,混合遗传算法(HGA)的优化曲线在经过 35 代后收敛而趋于稳定,得到最优解的值为 15.86。遗传算法(GA)的优化曲线经过 25 代进化后收敛而趋于稳定,得到最优解的值同样也为 15.86。从优化仿真实验结果来看,这两种智能算法对任务驱动的分组优化策略都是可行的。其中遗传算法的优化曲线较快地收敛,这验证了前文所叙述的遗传算法在求解具体问题时所表现出的全局搜索能力强而局部搜索能力较弱的结论。

当进化代数设为 150、种群为 30、交叉概率 $P_c=0.6$ 、变异概率 $P_m=0.02$ 时,HGA 算法(三角)和 GA 算法(方块)的优化曲线如图 3 所示。

如图 3 所示,随着进化代数的增加,最优解的值呈不断上升趋势,在经过 25 代进化后遗传算法(GA)的优化曲线收敛而趋于稳定,得到的最优值为 23.29;混合遗传算法(HGA)的优化曲线经过 35 代进化后收敛稳定,得到的最优值为 23.88。以上实验结果表明,运用混合遗传算法对任务驱动的分组优化效果明显更好。

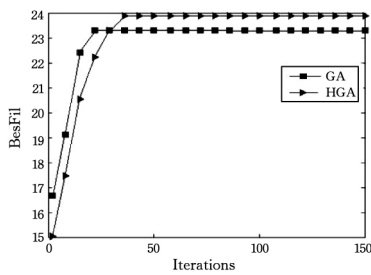


图3 进化曲线对比图(二)

运用混合遗传算法对任务驱动的分组进行优化后得到一个近似的最优解,即最优分组方案。在网络协作学习环境下的任务驱动教学中,以最优的分组方案进行协作学习小组划分,可以有效促使学习者更充分地进行讨论和交流,从而不会使得个别学习者对学习任务缺乏兴趣或者学习能力、理解力水平低而感到孤立^[16]。按最优的分组方案进行协作学习小组划分有助于学习者协作学习效率的提高;同时还有助于培养学习者的团队协作意识,从而为进一步开展网络协作学习营造良好的环境。

为了进一步验证混合遗传算法(HGA)和遗传算法(GA)对任务驱动分组优化的有效性,设定了4组不同的实验参数,并分别用HGA算法和GA算法对任务驱动的分组优化做10次测试,记录其中重复率较高的结果,对两种算法优化的最优解进行对比实验。实验结果如表2所列。

表2 不同参数下混合遗传算法和遗传算法仿真效果的对比

Gen	Pop	P_c	P_m	GHA/Fit	GA/Fit
50	10	0.5	0.01	7.98	7.98
100	20	0.6	0.03	15.86	15.27
150	30	0.7	0.05	23.88	16.65
200	40	0.8	0.07	31.52	17.12

如表2所列,基于不同实验参数分别用混合遗传算法(HGA)和遗传算法(GA)对任务驱动分组优化的效果进行对比实验,结果发现随着进化代数、种群数、交叉概率和变异概率的变化,混合遗传算法的优化效果明显要比遗传算法更胜一筹。

总之,从以上实验结果可以得出:基于混合遗传算法对任务驱动的分组优化是行之有效的。在实际网络学习环境下的基于任务驱动的分组教学中,运用混合遗传算法优化的最优分组方案进行小组划分,将有助于教学效率的提高;同时还有利于激发学习者学习的积极性,从而营造良好的网络协作学习环境。

结束语 人工智能在各领域的应用为具体问题的求解提供了智能化的解决方案。本研究将智能算法应用到教学领域中,构建了网络学习环境下的任务驱动分组优化问题的数学模型,并且利用混合遗传算法对任务驱动的分组优化进行了仿真实验。实验结果表明,利用混合遗传算法对网络环境下的任务驱动分组优化策略是行之有效的。在实际网络协作学习中,按优化的分组方案对学习者进行小组划分,对提高任务驱动的网络协作学习效率具有重大意义,同时还有利于激发学习者网络协作学习的积极性。本文在构建网络协作学习环

境中的任务驱动分组优化模型时,虽然考虑了学习者之间的个性特征差异和任务难易程度等因素,但实际网络协作学习分组过程中可能的影响因素远不止这些。因此,如何全面地考虑影响网络协作学习环境中学习者分组的影响因素并将其纳入优化模型中,进一步完善分组优化模型,对为网络协作学习提供更加智能化的分组支持服务的研究具有重要意义。

参考文献

- [1] HUXHAM M, LAND R. Assigning Students in Group Work Projects. Can We Do Better than Random? [J]. Innovations in Education & Training International, 2000, 37(1): 17-22.
- [2] 刘菊香. 基于模糊理论的网上协作学习学生分组系统的研究与实现[D]. 上海: 华东师范大学, 2006.
- [3] SUNG C T, SUNNY S J L. Learning through collaborative design: a learning strategy on the Internet [C] // Proceedings of the 2012 Frontiers in Education, 2001.
- [4] 马利东. 基于模糊聚类的网络合作学习研究及其算法实现[D]. 上海: 华东师范大学, 2003.
- [5] 陈瑶, 霍佳震. 整数编码的组群遗传算法在分组优化中的设计和应用[J]. 上海交通大学学报, 2013, 47(3): 472-478.
- [6] LIN Y T, HUANG Y M, CHENG S C. An automatic group composition system for composing collaborative learning groups using enhanced particle swarm optimization [J]. Computers & Education, 2010, 55(4): 1483-1493.
- [7] HWANG G J, YIN P Y, HWANG C W, et al. An enhanced genetic approach to composing cooperative learning groups for multiple grouping criteria [J]. Educational Technology & Society, 2008, 11(1): 148-167.
- [8] 胡慧, 何聚厚. 基于改进蚁群算法的协作学习分组研究[J]. 计算机工程与应用, 2014, 50(13): 137-141.
- [9] MORENO J, OVALLE D A, VICAR R M. A genetic algorithm approach for group formation in collaborative learning considering multiple student characteristics [J]. Computers & Education, 2012, 58(1): 560-569.
- [10] 王中华. 网络环境下协作学习的分组策略探究[J]. 软件导刊(教育技术), 2009(10): 44-45.
- [11] 应敏. 任务驱动的课堂合作学习智能分组系统的设计与开发研究[D]. 上海: 华东师范大学, 2007.
- [12] BEANE W E, LEMKE E A. Group variables influencing the transfer of conceptual behavior [J]. Journal of Educational Psychology, 1971, 62(3): 215-218.
- [13] 虞安波, 杨家本. 多背包问题的遗传算法求解[J]. 计算技术与自动化, 2002, 21(2): 59-63.
- [14] HOLLAND J H. Erratum: Genetic Algorithms and the Optimal Allocation of Trials [J]. Siam Journal on Computing, 1973, 2(2): 88-105.
- [15] 宋海生, 傅仁毅, 徐瑞松, 等. 求解多背包问题的混合遗传算法[J]. 计算机工程与应用, 2009, 45(20): 45-48.
- [16] YANG S J H. Context Aware Ubiquitous Learning Environments for Peer-to-Peer Collaborative Learning [J]. Educational Technology & Society, 2006, 9(1): 188-201.