

基于深度强化学习的自学习排课遗传算法研究

徐海涛, 程海燕, 童名文

引用本文

徐海涛, 程海燕, 童名文. [基于深度强化学习的自学习排课遗传算法研究](#)[J]. 计算机科学, 2024, 51(6A): 230600062-8.

XU Haitao, CHENG Haiyan, TONG Mingwen. [Study on Genetic Algorithm of Course Scheduling Based on Deep Reinforcement Learning](#) [J]. Computer Science, 2024, 51(6A): 230600062-8.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于改进遗传算法的家庭用电调度优化方法](#)

Scheduling Optimization Method for Household Electricity Consumption Based on Improved Genetic Algorithm

计算机科学, 2024, 51(6A): 230600096-6. <https://doi.org/10.11896/jsjcx.230600096>

[基于ART优化选择策略的遗传算法生成测试数据方法](#)

Method of Generating Test Data by Genetic Algorithm Based on ART Optimal Selection Strategy

计算机科学, 2024, 51(6): 95-103. <https://doi.org/10.11896/jsjcx.230100012>

[基于最大熵强化学习的最优渗透路径生成方法](#)

Optimal Penetration Path Generation Based on Maximum Entropy Reinforcement Learning

计算机科学, 2024, 51(3): 360-367. <https://doi.org/10.11896/jsjcx.221200104>

[基于遗传算法的生物启发频繁项集挖掘策略](#)

Bio-inspired Frequent Itemset Mining Strategy Based on Genetic Algorithm

计算机科学, 2023, 50(11A): 220700200-8. <https://doi.org/10.11896/jsjcx.220700200>

[基于GA-BP的圆形靶标圆心定位误差预测建模与补偿研究](#)

Study on Prediction Modeling and Compensation of Circular Target Center Positioning Error Based on GA-BP

计算机科学, 2023, 50(11A): 221100170-5. <https://doi.org/10.11896/jsjcx.221100170>

基于深度强化学习的自学习排课遗传算法研究

徐海涛 程海燕 童名文

华中师范大学人工智能教育学部 武汉 430079

(1332783819@qq.com)

摘要 排课是教学活动中一项常规而重要的事项,传统的人工排课方式费时费力,且容易出现错误,无法满足大规模排课的需求,而经典排课遗传算法存在收敛速度过快、排课效率随约束因素的增加而下降等问题。针对已有排课遗传算法存在的问题,提出一种基于深度强化学习的自学习排课遗传算法(GA-DRL)。GA-DRL算法利用Q-learning算法,实现了交叉参数和变异参数的自适应调整,增强了遗传算法的搜索能力,通过建立马尔可夫决策过程(MDP)的参数动态调整模型,对种群适应度函数进行状态集合的分析,实现对种群的整体性能的综合评价。同时将深度Q-网络算法(DQN)引入调度问题中,以解决排课中种群状态多、Q表数据量大的问题。实验结果表明,与经典排课遗传算法和改进的遗传算法相比,GA-DRL算法在正确率和寻优能力上有所提升。所提算法还可以应用于考场安排、电影院的排座和航空航线规划等问题。

关键词:排课问题;遗传算法;Q-学习;深度Q-网络

中图分类号 TP312;G473.4

Study on Genetic Algorithm of Course Scheduling Based on Deep Reinforcement Learning

XU Haitao, CHENG Haiyan and TONG Mingwen

Faculty of Artificial Intelligence in Education, Central China Normal University, Wuhan 430079, China

Abstract Course scheduling is a routine and important matter in teaching activities. The traditional manual course scheduling method is time-consuming and laborious, and prone to errors, which cannot meet the needs of large-scale course scheduling. However, the classical course scheduling genetic algorithm has problems such as too fast convergence speed and the efficiency of course scheduling decreases with the increase of constraint factors. Aiming at the problems of existing course scheduling genetic algorithms, a self-learning course scheduling genetic algorithm(GA-DRL) based on deep reinforcement learning is proposed. GA-DRL algorithm uses Q-learning algorithm to realize the adaptive adjustment of cross parameter and variation parameter, and enhances the searching ability of genetic algorithm. By establishing a dynamic parameter adjustment model of Markov decision process(MDP), the state set of fitness function is analyzed, and the overall performance of the population is evaluated comprehensively. At the same time, the deep Q-network algorithm(DQN) is introduced into the scheduling problem to solve the problem of multiple population states and large amount of Q-table data. Experimental results show that GA-DRL algorithm improves accuracy and optimization ability compared with the classical course scheduling genetic algorithm and improved genetic algorithm. The proposed algorithm can also be applied to problems such as examination room arrangement, cinema seating and airline route planning.

Keywords Scheduling questions, Genetic algorithm, Q-Learning, DQN

1 引言

排课是学校教学活动中一项常规而重要的工作,如何科学、高效地排课一直是教学活动安排中的一个重要问题。随着我国学校教育规模的不断扩大和教学方式的逐渐变化,传统人工排课的方式耗时、低效且缺乏科学性的问题日益凸显,已经不能满足现代教育的需求。

近年来,关于排课算法的研究受到了广泛关注,并且取得了丰富的研究成果,典型的算法有模拟退火算法、遗传算法和整数规划等^[1]。随着智能算法研究的发展,已有研究将启发式算法应用于解决排课问题,其中效果最好的是基于遗传算法(Genetic Algorithm, GA)的智能排课算法^[2]。但是目前这

类智能排课算法仍存在两个问题:1)该类算法有时收敛速度过快,影响了排课结果的准确性;2)随着引入的排课限制条件的增多,该类算法的执行效率大幅降低^[3-4]。产生这两个问题的主要原因是,现有的遗传算法的参数是人为设定或随机产生的,缺少对计算环境变化的适应性。因此,如何优化智能排课算法的参数需要做进一步的研究。

强化学习是一种机器学习方法,其基于环境的反馈而行动,通过不断与环境交互、试错,不断向积累回报提供最大的方向优化策略,最终使积累回报尽可能地最大化。强化学习的奖励很重要,其具有奖励导向性。研究表明,将强化学习引入遗传算法,能够通过揭示种群内部结构来搜寻到最优参数^[5-6]。目前已有学者利用强化学习对遗传算法参数进行研

究并取得了一定成果,例如,Zhang等^[7]将机器学习引入遗传算法中,提出了一种基于机器学习的多策略并行遗传算法,使用自主感知环境的强化学习,实现交叉概率参数优化,通过函数实例测试验证了改进遗传算法的优越性和稳定性。深度强化学习是强化学习和深度学习的结合,通过强化学习来定义问题本身和优化目标函数,通过深度学习来解决状态表示、策略表达等问题,采用深度神经网络来拟合强化学习中的值函数、策略,然后采用误差反向传播算法来优化目标函数^[8]。结构与强化学习类似,为解决状态-动作空间过大或连续状态-动作空间问题提供方便。Deng等^[9]针对复杂的数控铣削加工参数优化问题,提出了一种基于深度强化学习的数控铣削加工参数优化算法,通过实验验证,相对经验选取加工参数,提出的算法可以有效地优化参数,提高加工效率。

面向自动排课问题,基于深度强化学习和遗传算法的特点,设计了一种基于深度强化学习的自学习遗传算法(Self-learning Genetic Algorithm Based on Deep Reinforcement Learning,GA-DRL)。GA-DRL算法利用强化学习调整优化算法参数的特点来对遗传算法迭代过程中交叉、变异概率参数进行选择,保证种群的多样性,避免陷入局部最优,优化遗传算法,以解决当前智能排课算法排课结果准确性和排课效率对约束条件敏感的问题。

2 相关研究

排课是一类教学资源的最优分配问题,其本质上是带约束的0-1规划问题^[10],属于NP问题。目前常基于启发式算法解决这一问题,其中典型的算法有模拟退火算法^[11]、禁忌搜索算法^[12]、遗传算法^[13]、蚁群算法^[14]、粒子群算法^[15]、灰狼优化算法^[16]等。遗传算法是一种全局最优搜索改进算法,与其他的搜索改进算法相比,具有全局、并行搜索的特点,适用于求解排课问题。Ma等^[17]采用遗传算法设计排课系统,通过综合分析学生选择,生成最佳课程安排,实现自动排课,确保了课程安排的合理性,但缺乏对排课最优解收敛问题的研究。Chen等^[18]设计改进的遗传算法求解高中走班制下的排课问题,通过制定有针对性的适应度函数,进行两次遗传操作,最终得到较合理的排课结果,但同样缺乏对算法收敛性的研究。由于排课问题的复杂性及遗传算法自身的缺陷,因此需要对遗传算法进行优化,如早熟、处理规模小^[19]。Ma等^[20]提出了自适应的交叉概率和变异概率设计方案,有效防止陷于局部最优解,提高了全局搜索能力。事实上,遗传算法的性能与关键参数的选择有很大关系,但对于如何确定遗传算法的参数以改善计算效能以及解决问题的效率,目前学术界的研究还相对较少。

近年来,随着机器学习研究广泛而深入的开展,许多研究人员尝试将强化学习应用到启发式算法的参数动态调节中,以实现对最优化问题的求解。针对遗传算法,有研究发现强化学习具有发现群体的内在结构,从而找到最优参数的优点^[21]。此外,还有一些研究应用强化学习对最优算法的参数进行调整,并将其用于产品制造^[22]、车间调度^[23]、电梯调度^[24]等过程,以提高启发式算法解决问题的精度和效率。但是,目前尚缺少将强化学习应用于智能排课算法的相关研究。

本文第2章简要介绍了排课问题的相关研究;第3章详

细介绍了所提出的GA-DRL的模型框架及算法的实现;第4章给出了算法参数设置过程,以及与其他两种排课算法的对比实验,并对结果进行了分析阐述;最后总结全文并展望未来。

3 基于深度强化学习的智能排课算法设计

3.1 问题描述

排课是给多个教学任务安排互相不冲突的上课时间和教室的过程,其中教学任务是由教师、学生、课程组成的三元组。排课中可能会出现多种冲突,因此必须满足一些硬性约束条件才能避免排课中的冲突。常见硬性约束条件有:1)同一时段,同一个教室至多能开设一门课;2)同一时段,同一名教师至多能开设一门课程;3)同一时段,同一班级至多能开设一门课程。

排课除满足硬性约束条件以外,也应考虑学校、教师和学生需求,以达到资源利用率最大、教学效果最优的目标。例如,教师可以根据学生的学习进度来调整课程安排,以确保教学效果。再如排课应考虑到教师的上课意愿,以确保教师能按时授课。因此,排课还需要满足一些软性约束条件。典型的软性约束条件包括:1)课程应均匀分布在一个星期的所有开课时段;2)特殊课程安排在特殊时段,例如下午第一节最好安排自习课;3)一个班级在同一天内不要安排相同的课程;4)教师的上课时段需求。

3.2 数学模型

课程表是一个五元组(课程、班级、教师、教室、时段),分别用 L, C, H, R, T 表示。每个元组是一个集合,具体如下:

$$L = \{l_1, l_2, l_3, \dots, l_L\};$$

$$C = \{c_1, c_2, c_3, \dots, c_C\};$$

$$H = \{h_1, h_2, h_3, \dots, h_H\};$$

$$R = \{r_1, r_2, r_3, \dots, r_R\};$$

$$T = \{t_1, t_2, t_3, \dots, t_T\};$$

用 $T^i = \{t_1^i, t_2^i, t_3^i, \dots, t_T^i\}$ 表示第 i 天的时段集合。

因此,排课问题的常见硬性约束条件可以表示为:

1)在 i 时段的同一个教室里,最多只能开一门课程,数学模型表示如下:

$$\forall i \in T, \forall j \in R, \left(\bigcap_{k=1}^R C_{k(i,j)} \right) \leq 1 \quad (1)$$

2)在 i 的时段中,一位老师只能在 R_j 的一个课堂中讲授一门课,数学模型表示如下:

$$\forall i \in T, \bigcap_{j=1}^R H(i, R_j) = \emptyset \quad (2)$$

3)在 i 的时段里,一个班级只能有一门课被安排在 R_j 的一个教室里,数学模型表示如下:

$$\forall i \in T, \bigcap_{j=1}^R C(i, R_j) = \emptyset \quad (3)$$

排课问题的常见软性约束条件可以表示为:

在某一天任意 i 的时段里,一个班级不能安排相同的课程,用 k 表示星期几,数学模型表示如下:

$$\forall i \in L, \bigcup_{k=1}^5 T^k(i, C_j) = \emptyset \quad (4)$$

因此,可以将排课问题描述为一个单目标最优化模型。

决策变量为: $X(l, c, h, r, t)$,表示在 t 的时段, h 教师在 r 教室给 c 班级教授 l 课程。

目标函数为:

$$\min = \sum_{i=1}^5 \sum_{l=1}^L \text{card}(\bigcap_{c=1}^C X_{i \in T^l}(l, c, h, r, t)) \quad (5)$$

约束条件为式(1)一式(3)。

3.3 排课遗传算法设计

目前,遗传算法是解决带约束 0-1 规划问题的有效算法^[25]。遗传算法是基于模拟生物学家达尔文提出的“适者生存,优胜劣汰”的生物进化论演变而来的一种启发式搜索算法,能够在搜索时自动地获得并累积关于搜索空间的相关知识,通过执行选择、交叉和变异操作,使问题朝着更好的方向发展,以获得最优解。因为排课是一类带约束的 0-1 规划问题,所以可以采用遗传算法来解决。

3.3.1 染色体表示和种群的初始化

染色体是由各种基因构成,每一条代表一种课程安排,例如,某个基因可以表示某个教学任务安排在某个教室,而某一条染色体才可以表示为问题求解的最优解。在排班之前,采用十进制方法,对课程号、班级号及教师号进行排班,同时由计算机自动排班表生成班级号和课时号。其中时间编号是按照每周和每日的时间序列进行编码的,本文算法设置每个星期共五天的课程,并将其划分为每日六次,每次两节课,如 16 表示的是周日的第 11~12 节课,以此类推。编码表示为“课程单元编码+班级编码+教师编码+教室编码+时间编码”,如表 1 所列,例如教师工号 11,课程《计算机网络》(编号 100044),若将此课程安排到周二第 1 节和第 2 节课,授课班级为 3 班(编码为 3),在 105 教室(编码为 5)上课,则此基因编码为 100044311521,其中末尾的 21 编码代表周二第 1~2 节课。

表 1 基因编码

Table 1 Gene coding

课程号	班级号	教师号	教室	时间
6 位	1 位	2 位	1 位	2 位

3.3.2 遗传操作

本文选择排班方案冲突值这一目标函数为适应性函数,而冲突则涉及多个方面,如课堂冲突、班级冲突、教师冲突、课程分配的和諧度问题等,本文把自适应函数设计为对每一个冲突的数值进行加权得到的结果。本文采用一种基于随机遍历法的选择策略,从父代中挑选出一批具有高适应度(冲突值较小)的个体,并将其作为后代,然后通过对选择出的优等个体进行杂交和变异,得到新的优等个体。

自适应函数定义为:

$$f = \sum_{i=1}^n w_i f_i \quad (6)$$

其中, w_i 为每个冲突因素的权重,它的值可以按照实际情况来确定, f_i 为各个冲突因素, f 为排课问题的目标函数。

1) 教室冲突:同一时间,在同一个教室里,最多只能开一门课程,若出现一次冲突,则 $f_1 = f_1 + 1$ 。

2) 班级冲突:同一时间,一名老师只能出现在一个课堂上讲授一门课程,若出现一次冲突,则 $f_2 = f_2 + 1$ 。

3) 教师冲突:同一时间,一个班级至多只可以开设一门课程,若出现一次冲突,则 $f_3 = f_3 + 1$ 。

4) 课程时间离散均匀度:一个班级在同一天内不能安排相同的课程,若出现一次冲突,则 $f_4 = f_4 + 1$ 。

此处用“冲突”来描述适应性,当“冲突”较大时,相应的适应性就较低;反之,当“冲突”较小时,适应性就较高。

3.4 Q-learning 算法设计

Q-learning 算法将蒙特卡罗算法与动态规划算法的思想相融合,因此被认为是最有效的强化学习算法。在 Q-learning 算法中,马尔可夫决策模型(MDP)的构建是一个关键问题。MDP 模型一般由 4 个要素组成,即状态集、奖励函数、动作集和策略函数。Q-learning 算法使用一张 Q-table 来记录每一种操作的行为值,因此当操作或状态空间越大,所需的存储就越多。由于排课问题所选择的状态集过大且不确定,因此本文引入了 DQN 算法。DQN 算法的关键之处在于用了一个人工神经网络 $q(s, a; \omega)$, $s \in S, a \in A$ 来代替 Q-table,即动作价值函数。网络的输入为状态信息,输出为每个动作的价值,如图 1 所示。DQN 算法具有较强鲁棒性,能够很好地处理连续和不确定性的状态集合问题。DQN 算法将 3 个因素结合起来,即以往的经验状态、选择行为以及所处的环境所给予的回报,对 NN 表进行了更新。NN 表的更新表达式为:

$$Q(s, a) \leftarrow Q(s, a) + \alpha [r + \gamma \max_{a'} Q(s', a') - Q(s, a)] \quad (7)$$

其中, $Q(s, a)$ 为智能体处于状态 s 和动作 a 时的动作价值; $Q(s', a')$ 为智能体处于状态 s' 和动作 a' 时的动作价值; r 为即时奖励函数; α 为学习率; γ 为折扣率。

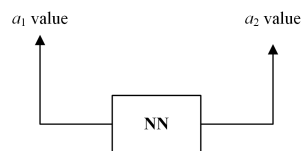


图 1 人工神经网络

Fig. 1 Artificial neural network

3.4.1 状态集设计

排课最优化问题中,目标函数是使冲突值为 0,在该算法中,冲突值就是群体的适应度函数。为了对群体的表现做出一个全面的评价,本文选择了两个指标,即群体的平均适合度和最佳的个体适合度,从而对状态集进行划分。设 $G(x_i^t)$ 是第 t 次迭代时第 i 个个体的适应函数,那么在第 t 次迭代次数下,该群体的平均适应函数 G^t 可以被确定为:

$$G^t = \sum_{i=1}^N G(x_i^t) / N \quad (8)$$

其中, x_i^t 表示第 t 次迭代时的第 i 个个体, N 为种群个体数量。

种群在第 t 代最小个体的适应度函数 M^t 定义为:

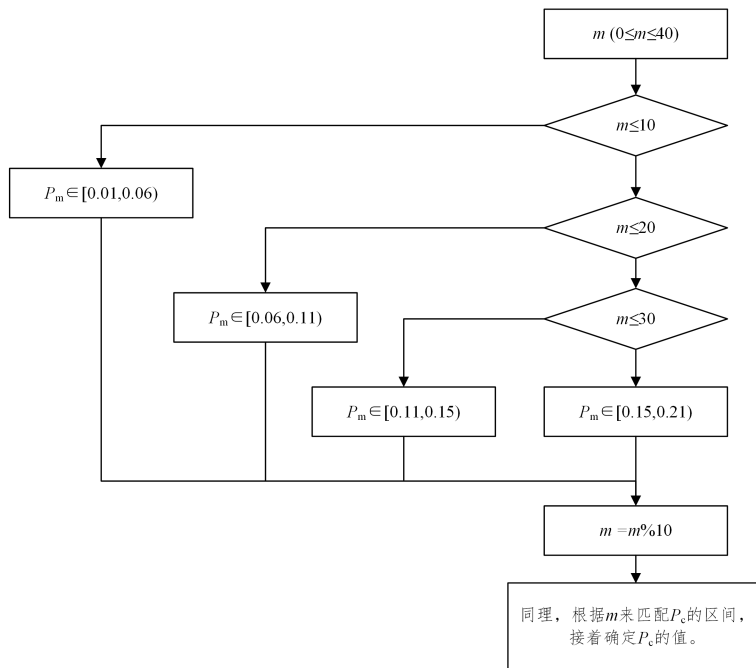
$$M^t = \min G(x_i^t) \quad (9)$$

则种群状态可由向量 S 表示为:

$$S = (G^t, M^t) \quad (10)$$

3.4.2 动作集和奖励函数的设置

在每一次迭代中,智能体都会选择不同的动作,以获得适当的交叉概率 P_c 和变异概率 P_m 。在交叉概率方面,一般选取 0.4~0.9 的区间,以 0.05 为间隔将其分成 10 个区间;在变异概率方面,一般取值在 0.01~0.21 之间,本文将其分成 4 个区间,间隔为 0.05,则本文提出的 MDP 模型中的动作集为交叉概率集和变异概率集的笛卡尔积,共 40 个区间。遗传算法根据动作值 m 来确定 P_c 和 P_m 的值,如图 2 所示。

图2 P_c 和 P_m 确定的流程图Fig. 2 Flow chart of determining P_c and P_m

接下来,选择一个适当的奖励函数,对交叉概率以及变异概率选择进行评估,本文定义奖励函数 r 为:

$$r = \frac{\sum_{i=1}^N G(x_i^t) - \sum_{i=1}^N G(x_i^{t-1})}{\sum_{i=1}^N G(x_i^{t-1})} \quad (11)$$

其中, $G(x_i^{t-1})$ 表示上一代第 i 个个体的适应度函数,由计算式可以看出,如果目前的个体的平均适应度函数比上一代要好,即当 r 是正值时,智能体得到了嘉奖,说明当前的 P_c 和 P_m 是有效的,否则无效。

3.4.3 动作选择策略

选择强化学习是最基础的工作,使用频率最高的 ϵ -greedy 策略为动作选择策略, ϵ -greedy 策略在使用与寻找之间取得了一种平衡,其中选取最大的行为值函数对应的行为,并以一定的概率选择其他非最优行为。

3.5 基于深度强化学习的自学习排课遗传算法设计

基本遗传算法的核心是交叉与变异运算,其具有非常关键的影响,而交叉与变异运算的主要参量是 P_c 和 P_m 。在遗传算法基因演算的过程中,若 P_c 和 P_m 取值过大,则会导致计算速度缓慢,难以寻求最优解;而若 P_c 和 P_m 的值过小则很难获得新的个体。强化学习的优势在于,它能够有效地发现种群的内部结构,通过深入挖掘个体间的交互信息,并利用环境状态变化来确定最优的参数,实现最优的决策。此外,还可以通过评估反馈,不断调整参数,以获得更好的结果。同时,针对操作状态集过大的问题,本文引入 DQN 算法来动态地调整遗传算法中的 P_c 和 P_m 值,以期望改善该算法的全局寻优能力。

DQN 深度强化学习调整 P_c 和 P_m 大致可分为 4 个步骤:首先,智能体通过收集环境中的状态 s ,对其进行评估,计算出适应度,以确定当前的环境状态;其次,采用选择的策略 ϵ -greedy,智能体根据神经网络计算出来的相应值选择动作 a ;紧接着,根据当前状态 s 和策略 ϵ -greedy 选出来

的动作,执行一次遗传进化操作,这时,种群的状态转变成 s' ,并对其进行反馈,以及给出 reward,若所得 reward 为正,那么遗传算法的动作选择会变得更强,若所得 reward 为负,那么它的选择会变得更弱;最后,将操作 s 、选择的动作 a 、奖励 r 及操作 s' 存储到记忆池中,再选择性地 DQN 学习,然后智能体重新选择动作执行。在此基础上,根据以往的训练结果,深度强化这一个过程被激活,因此 P_c 和 P_m 的动作选择将得到优化。基于以上的一系列分析,GA-DRL 算法的伪代码如算法 1 所示,完整 GA-DRL 算法流程图如图 3 所示。

算法 1 基于深度强化学习的自学习遗传算法

输入: teachers.csv, courses.csv, training plans.csv

输出: course arrangement for each class.xls

1. schedules, courses, teachers \leftarrow init("teachers.csv", "courses.csv", "training plans.csv");
2. init GeneticPopulation, DQN;
3. $s \leftarrow$ reset();
4. while $n <$ MaxEvolution
5. $a \leftarrow$ choose_action(s);
6. $s', r, result \leftarrow$ step(s, a);
7. NN \leftarrow store_transition(s, a, s', r);
8. if $n <$ 200 and $n \% 5 = 0$ then
9. DQN.learn();
10. $s = s'$;
11. if result \neq None then
12. end while;
13. end while

在 GA-DRL 算法的流程中,每次遗传算法的迭代过程中,智能体都会根据策略 ϵ -greedy 选择动作,即选择交叉概率 P_c 和变异概率 P_m 。本文算法基于遗传算法的内部状态,通过深度强化学习来动态调节遗传算法的变异概率和交叉概率参数的这一过程,即实现遗传算法自学习的过程。

GA-DRL 算法的计算复杂度主要来源于遗传算法中选择、交叉、变异等操作及学习神经网络的更新操作。因此 GA-DRL 的总计算时间复杂度(T)可以表示为:

$$T=O((T * N)+(B * E * N)) \quad (12)$$

其中, T 是迭代次数, N 是种群大小, B 是每次更新 NN 表时的样本数量, E 是训练的迭代次数, N 是神经网络的参数个数。

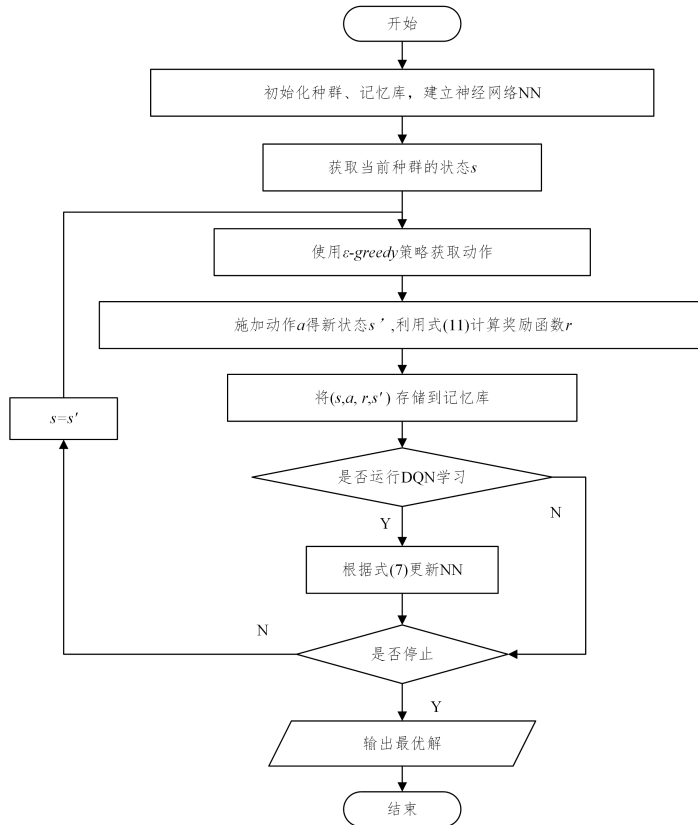


图3 GA-DRL算法的流程图

Fig. 3 Flow chart of GA-DRL algorithm

4 实验结果与分析

4.1 实验设计

在实验中,设定排课数据为:5个班级,23个教师,10个教室,15门课程,上课时间为星期一到星期五,每天6次课程。将GA-DRL算法和传统的遗传算法作为基线,同时与文献[20]中的改进遗传算法进行比较。

为了对本文提出的排课算法的精度和性能进行评价,本文提出了一些重要的评价指标:可行性、排课成功所需迭代次数、成功概率、最优适应度、收敛速度、稳定性。其中,可行性指算法生成的解是否满足所有的约束条件;排课成功所需迭代次数指算法达到最优解所需要的迭代次数;成功概率指在相同测试样本的条件下,排课成功次数与总测试数的比值;最优适应度指算法最终生成的种群中适应度最小的个体的适应度,即冲突值,该指标直接反映了算法质量的好坏,最优解适应度越小,算法的质量越好,反之,算法的质量越差;收敛速度指最优适应度随进化代数下降的速度;稳定性指在相同测试样本的条件下,算法在多次运行中是否产生一致的结果。

在满足可行性条件的前提下,随着任务规模的扩大,考虑评估排课算法的迭代次数和成功概率。高迭代次数可以提高算法的精度,而较低的迭代次数可能导致精度下降。此外,成功概率的提高也与算法的精度正相关,而成功概率的降低可能导致精度减弱。同时,考虑排课算法的收敛速度和稳定

性,较快的收敛速度表明算法的性能较强,而较慢的收敛速度可能暗示算法的性能较弱。对于稳定性,较高的稳定性意味着算法的性能较强,而较低的稳定性可能导致算法性能下降。

4.2 实验环境

为了对本文提出的自动排课算法的有效性进行测试,选取了普通的遗传算法(记为GA)和文献[20]中的改进遗传算法(记为改进遗传算法^[20])与其进行比较,三者使用相同的染色体编码方案并具有相同的适应度函数。该实验是在一台2.50GHz处理器和16GB RAM,操作系统为Win11的计算机上进行的,编程工具采用PyCharm Community Edition 2022,编程语言采用python。此外,GA-DRL算法中相关参数设置为:学习率 $\alpha=0.01$,折扣率 $\gamma=0.9$,贪婪率 $\epsilon=0.9$,记忆库的容量 $MS=500$,更新Q现实网络参数的步骤 $RTI=200$,每次从记忆库中取的样本数量 $BS=32$,交叉和变异概率利用DQN动态调整。

4.3 实验结果

为了测试和评估GA-DRL算法的准确性,分别对培养任务个数为45,50,55,60,65的5种情况各进行了10次模拟测试并记录,将它们的平均值作为实验数据结果。培养任务数量为65的某一次测试结果如图4所示。此外,为了检测GA-DRL算法的性能,本文将GA-DRL算法与GA和改进遗传算法^[20]两种算法就排课时间、排课迭代次数、成功率进行了比较,结果如表2所列。

1 班						2 班					
周/节	Mon	Tue	Wed	Thu	Fri	周/节	Mon	Tue	Wed	Thu	Fri
1			计算机网络 地点:103 教师:孙文			1	计算机网络 地点:109 教师:周友国		数据结构地点:106 教师:祝毅勇	教学设计 地点:102 教师:冯艺仁	
2			计算机应用 基础 地点:103 教师:赵思斌	计算机网络 实验 地点:101 教师:赵国		2		人工智能 基础 地点:106 教师:王宏大			
3				数据结构 地点:107 教师:祝毅勇		3	嵌入式原理 地点:109 教师:冯桥				
4	计算机网络 地点:106 教师:周友国	电路设计 实践 地点:110 教师:董仁兴	计算机组成 原理地点:109 教师:赵国健	课程研究 地点:105 教师:郑军锦		4			计算机 网络实验 地点:101 教师:赵国	计算机 网络实验 地点:101 教师:项友志振	
5				电路设计 实践 地点:105 教师:董仁兴	数据库 基础原理 地点:104 教师:王磊宏	5			编译原理 地点:103 教师:周友国	大学 C 语言 地点:110 教师:祝毅勇	数据库基础 原理 地点:105 教师:李兴
6	编译原理 地点:106 教师:周友国	教学设计 地点:107 教师:王磊宏		并行程序 设计基础 地点:107 教师:祝桥		6	电路设计 实践 地点:102 教师:赵宏生			电路设计 实践 地点:108 教师:赵宏生	编译原理 地点:110 教师:项峰炎
3 班						4 班					
周/节	Mon	Tue	Wed	Thu	Fri	周/节	Mon	Tue	Wed	Thu	Fri
1	计算机应用 基础 地点:107 教师:赵思斌		数据库 基础原理地点:109 教师:赵荣正		数据结构 地点:109 教师:祝毅勇	1		课程研究 地点:107 教师:董南振			
2	教学设计 地点:103 教师:赵荣正				大学 C 语言 地点:103 教师:项友志	2	嵌入式原理 地点:107 教师:冯桥	计算机 应用基础 地点:102 教师:赵宏生			机器学习 地点:107 教师:李艺祥
3			并行程序 设计基础 地点:108 教师:祝桥		电路设计 实践 地点:108 教师:董友裕	3			计算机网络 实验 地点:110 教师:赵国	大学 C 语言 地点:102 教师:董南振	
4	编译原理 地点:101 教师:项健德		课程研究 地点:103 教师:董南振	计算机网络 地点:108 教师:周友国	计算机组成 原理地点:104 教师:冯桥	4	电路设计 实践 地点:107 教师:董仁兴			数据库 基础原理 地点:110 教师:李兴	
5		人工智能 基础 地点:102 教师:冯艺仁				5		计算机组成 原理 地点:101 教师:郑军锦	教学设计 地点:102 教师:王磊宏		
6			机器学习 地点:102 教师:项健德			6	并行程序 设计基础 地点:109 教师:祝桥	计算机网络 地点:102 教师:孙文		人工智能 基础 地点:103 教师:赵荣正	
5 班											
周/节	Mon	Tue	Wed	Thu	Fri	周/节	Mon	Tue	Wed	Thu	Fri
1			机器学习 地点:110 教师:祝桥			1					
2	计算机组成原理 地点:110 教师:赵国健		教学设计 地点:105 教师:冯艺仁		课程研究 地点:104 教师:董南振	2					
3			数据库基础原理 地点:101 教师:李兴			3			数据结构 地点:106 教师:李兴	数据结构 地点:107 教师:李兴	
4						4					
5	并行程序设计基础 地点:104 教师:郑宏文			计算机网络实验 地点:109 教师:项友志	人工智能基础 地点:108 教师:王磊宏	5				机器学习 地点:103 教师:项健德	
6				计算机应用基础 地点:104 教师:赵思斌	编译原理 地点:102 教师:项峰炎	6					

图 4 GA-DRL 算法运行结果
Fig. 4 GA-DRL algorithm running results

表2 3种算法的结果
Table 2 Results of three algorithms

培养任务/个	GA			改进遗传算法 ^[20]			GA-DRL		
	time/s	iteration/代	success rate	time/s	iteration/代	success rate	time/s	iteration/代	success rate
45	0.449	7	1	0.48	6	1	1.929	7	1
50	20.446	384	0.7	2.966	38	1	19.286	133	1
55	31.244	509	0.4	25.22	286	1	43.093	275	1
60	107.08	1530	0.1	57.027	540	0.8	66.205	347	0.9
65			0	119.25	1041	0.6	166.59	853	0.8

4.3.1 算法精度

对图4进行检测,可以依次对同一时间的课程进行对比检查,首先,可以发现没有班级在同一时间开设两门课程;其次,可以发现在同一时间没有存在相同教室被使用的情况;最后,可以发现在同一时间没有相同的教师在上课。除此之外,对GA-DRL算法运行成功的十余次结果进行检测,发现都是正确的,因此可以得出结论:GA-DRL算法的结果是可靠的。

由表2可以看出,在低培养任务量下,GA-DRL算法排课时间不如GA算法和改进遗传算法^[20];随着培养任务数量的增加,GA算法求解时间大幅增加,直至到一定时间内求不出结果;随着培养任务数量的增加,改进遗传算法^[20]和GA-DRL算法排课所需时间短于GA算法,很大程度提升了排课的效率,而GA-DRL算法在迭代过程中需要花费额外时间进行记忆池更新学习及重复从NN表中选择合适的变异概率和

交叉概率参数,因此该算法相对改进遗传算法^[20]用时稍长;但从迭代次数上看,随着培养任务量的增加,GA-DRL算法能在更少迭代次数内得到排课结果。此外,随着培养任务量的增加,GA-DRL算法的排课测试成功率依然很高。

4.3.2 算法性能

为进一步验证GA-DRL的性能,本文选取了部分培养任务规模的最优适应度值和运行成功所需迭代次数,对改进遗传算法^[20]和GA-DRL两种算法测试成功的数据进行统计分析,将它们的平均值作为实验数据结果,并每进化100代记录1次该种群的最优适应度值,最终将这些数据的平均值作为适应度值方面和运行所需迭代次数的实验结果,在不同规模($S=55,60,65$)下,两种算法最优个体适应度值和迭代次数关系的结果如图5所示,两种算法运行成功所需迭代次数的结果如图6所示。

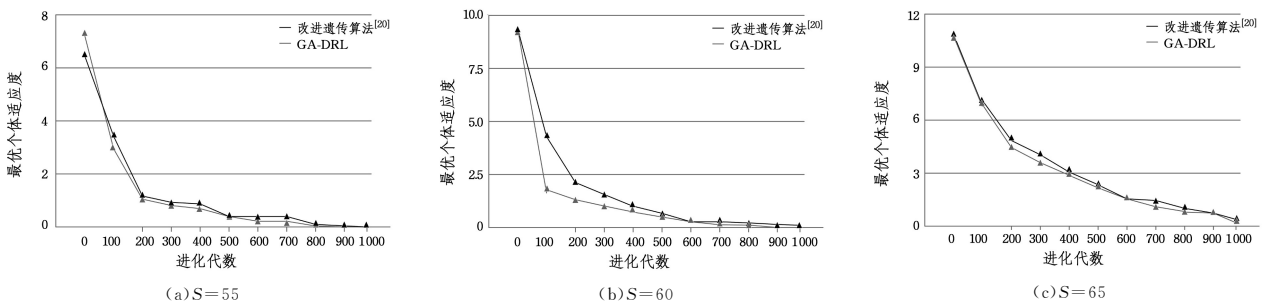


图5 培养任务数量 $S=55,60,65$ 时,改进遗传算法^[20]和GA-DRL算法的收敛曲线

Fig. 5 Convergence curve of improved genetic algorithm^[20] and GA-DRL algorithm when $S=55,60,65$

从图5的寻优曲线可以看出,对于培养任务数量 $S=55,60,65$,本文提出的基于深度强化学习的自学习遗传算法

(GA-DRL)的寻优性能优于改进遗传算法^[20],收敛速度略胜于改进遗传算法^[20]。

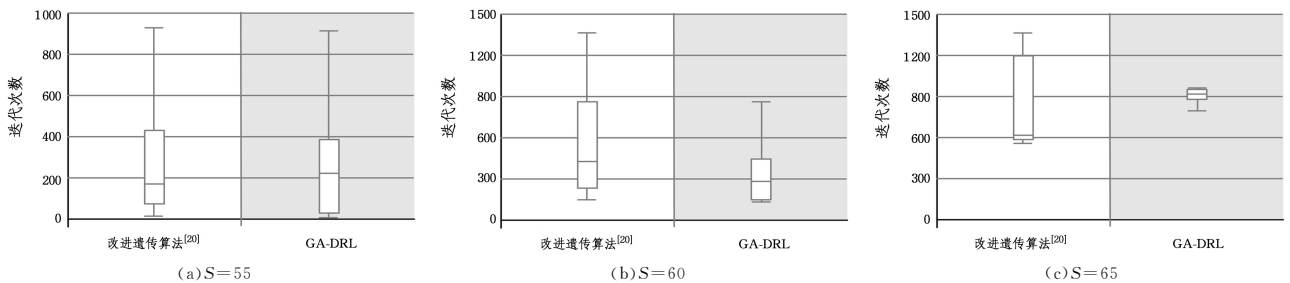


图6 培养任务数量 $S=55,60,65$ 时,改进遗传算法^[20]和GA-DRL算法运行所需迭代次数的箱型图

Fig. 6 Box diagram of the number of iterations required by the operation of improved genetic algorithm^[20] and GA-DRL when $S=55,60,65$

此外,由图6的箱型图可以看出,对于培养任务数量 $S=55,60,65$,GA-DRL算法的中值近似等于改进遗传算法^[20]。当 $S=60$ 时,GA-DRL算法的中值明显小于改进遗传算法^[20];而在3个规模下,GA-DRL算法的波动范围都明显小于改进遗传算法^[20]。

综上所述,本文提出的GA-DRL算法在精度和寻优性能方面胜于GA算法,略胜于改进遗传算法^[20],适合应用到排课问题中。

结束语 本文为自动求解排课问题提出了一种基于深度强化学习的自学习排课遗传算法DRL-GA,对不同规模培养任务量进行实验研究。实验结果表明,该算法具有较好的搜索性能,能够自动完成排课过程,排课的精度和寻优性能均优于经典的遗传算法,能在保证排课结果无冲突的前提下,显著减少排课的迭代次数,从而可以为高校的教学工作提供更加准确、可靠的数据支持。该算法还可以应用于考场安排、船舶

调度和航空航线规划等问题。另外,该算法设计仅仅是在设置了3个硬性约束条件和1个软性约束条件下得到的满意解,缺少对课表满意度的指标优化。下一步研究还需要增加更多的软性约束条件,在保证课表准确性的前提下满足更多的软性约束条件,排出更符合实际情况的课表。

参 考 文 献

- [1] WANG W H, LI W Q. Timetabling algorithm of high school optional class system based on improved genetic algorithm[J]. Journal of Zhejiang University of Technology, 2016, 44(6): 601-607, 670.
- [2] ZHANG Z F. Design of an intelligent timetabling algorithm[J]. Electronic Design Engineering, 2022, 30(9): 134-138.
- [3] JIANG J, BAI S X. Improvement of Genetic Algorithm and Its Application in Course Scheduling [J]. Journal of Nanchang University(Natural Science), 2018, 42(4): 388-392.
- [4] SUN Y, HU J H. Course Schedule System Based on Genetic-Ant Colony Hybrid Algorithm[J]. Computer Systems & Applications, 2019, 28(2): 81-86.
- [5] CAO Z C, LIN C R, ZHOU C. A knowledge-based cuckoo search algorithm to schedule a flexible job shop with sequencing flexibility[J]. IEEE Transactions on Automation Science and Engineering, 2019, 18(1): 56-69.
- [6] JIAN Z, LIU S, ZHOU M, et al. Modified cuckoo search algorithm to solve economic power dispatch optimization problems [J]. IEEE/CAA Journal of Automatica Sinica, 2018, 5(4): 794-806.
- [7] ZHANG Y, ZHONG H C, ZHANG C J, et al. Multi-strategy parallel genetic algorithm based on machine learning[J]. Computer Integrated Manufacturing Systems, 2021, 27(10): 2921-2928.
- [8] YANG W, ZHI B C, WU Z R, et al. Review of Reinforcement Learning for Combinatorial Optimization Problem[J]. Journal of Frontiers of Computer Science & Technology, 2022, 16(2): 261.
- [9] DENG Q L, LU J, CHEN Y H, et al. Optimization method of CNC milling parameters based on deep reinforcement learning [J]. Journal of Zhejiang University (Engineering), 2022, 56(11): 2145-2155.
- [10] WANG Y C, ZHOU C H, CHEN C X, et al. Extensible Intelligent Scheduling System Based on Two Fork Knowledge Tree Reasoning[J]. Computer Engineering and Applications, 2021, 57(5): 251-257.
- [11] SONG T, CHEN M, WU C, et al. Automated course arrangement algorithm based on multi-class iterated local search[J]. Comput Appl, 2019, 39(6): 1760-1765.
- [12] ZHANG Y. Design and application of course scheduling system based on tabu search method[J]. Electronic Design Engineering, 2018, 26(16): 40-44.
- [13] CHEN L, XIU W. Solution to course-timetabling problem of mobile learning system based on improved genetic algorithm[J]. Comput Eng Appl, 2019, 55(6): 224-230.
- [14] JIANG Z F, QIN H, PEI-PEIL U, et al. Research on the application of ant colony algorithms to course scheduling in colleges and universities[J]. Modern Computer, 2019, 10(45): 23-33.
- [15] LUO Y, CHEN Z, SCIENCE F O. Optimization of university course scheduling problem based on improved pso algorithm[J]. Computer Applications and Software, 2018, 35(6): 241-247, 303.
- [16] LONG Z, AI J, ZOU H, et al. Network traffic predicting model based on improved grey wolf optimization algorithm[J]. Application Research of Computers, 2018, 35(6): 1845-1848.
- [17] MA Y F, ZHANG H N, JIE S. Research and Implementation of Genetic Algorithm in Curriculum Arrangement System in Colleges and Universities [J]. Computer Systems & Applications, 2014, 23(5): 112-115.
- [18] CHEN L, XIU W. Solution to course-timetabling problem of mobile learning system based on improved genetic algorithm[J]. Comput Eng Appl, 2019, 55(6): 224-230.
- [19] ZHANG H, LUO Y, ZHANG L, et al. Cultivated land change forecast based on genetic algorithm and least squares support vector machines[J]. Transactions of the Chinese Society of Agricultural Engineering, 2009, 25(7): 226-231.
- [20] MA X Z, LI F R. Application Research of course scheduling System based on improved Genetic Algorithm[J]. Journal of Qingdao University(Engineering Technology Edition), 2020, 35(3): 15-20, 55.
- [21] SHAHRABI J, ADIBI M A, MAHOOTCHI M. A reinforcement learning approach to parameter estimation in dynamic job shop scheduling [J]. Computers & Industrial Engineering, 2017, 110(8): 75-82.
- [22] BI W J, ZHOU Y B. Research on inventory control and dynamic pricing of fresh produce based on deep reinforcement learning [J]. Application Research of Computers, 2022, 39(9): 2660-2664.
- [23] LI B S, YEC M. Job shop scheduling problem based on deep reinforcement learning [J]. Computer Engineering and Applications, 2021, 23: 248-254.
- [24] LIU Y, ZHANG C, LI T. Modeling and application of reinforcement learning A3C in elevator scheduling algorithm[J]. Computer Engineering and Design, 2022, 43(1): 196-202.
- [25] FAN Y. Genetic algorithm for solving linear bilevel programming with interval coefficients[J]. Journal of Computer Applications, 2014, 34(1): 185.



XU Haitao, born in 2000, undergraduate. His main research interests include machine learning and optimal algorithm design.



TONG Mingwen, born in 1975, doctor, professor. His main research interests include adaptive learning theory and method, knowledge modeling.