

求解多目标流水车间调度 Pareto 最优解的遗传强化算法^①



刘宇, 陈永灿, 周艳平

(青岛科技大学 信息科学技术学院, 青岛 266061)

通信作者: 周艳平, E-mail: ypzhou@qust.edu.cn

摘要: 针对多目标流水车间调度 Pareto 最优问题, 本文建立了以最大完工时间和最大延迟时间为优化目标的多目标流水车间调度问题模型, 并设计了一种基于 Q-learning 的遗传强化学习算法求解该问题的 Pareto 最优解. 该算法引入状态变量和动作变量, 通过 Q-learning 算法获得初始种群, 以提高初始解质量. 在算法进化过程中, 利用 Q 表指导变异操作, 扩大局部搜索范围. 采用 Pareto 快速非支配排序以及拥挤度计算提高解的质量以及多样性, 逐步获得 Pareto 最优解. 通过与遗传算法、NSGA-II 算法和 Q-learning 算法进行对比实验, 验证了改进后的遗传强化算法在求解多目标流水车间调度问题 Pareto 最优解的有效性.

关键词: 多目标流水车间调度; Q-learning; 遗传算法; Pareto

引用格式: 刘宇, 陈永灿, 周艳平. 求解多目标流水车间调度 Pareto 最优解的遗传强化算法. 计算机系统应用, 2024, 33(2): 239-245. <http://www.c-s-a.org.cn/1003-3254/9428.html>

Genetic Reinforcement Algorithm for Solving Pareto Optimal Solutions for Multi-objective Flow Shop Scheduling

LIU Yu, CHEN Yong-Can, ZHOU Yan-Ping

(School of Information Science & Technology, Qingdao University of Science and Technology, Qingdao 266061, China)

Abstract: Aiming at the Pareto optimal problem for multi-objective flow shop scheduling, this study builds a multi-objective flow shop scheduling problem model with maximum completion time and maximum delay time as the optimization objectives. Meanwhile, the study designs a genetic reinforcement learning algorithm based on Q-learning for the Pareto optimal solution of the problem. The algorithm introduces state variables and action variables and obtains the initial population by Q-learning algorithm to improve the initial solution quality. During the evolution of the algorithm, the Q-table is applied to guide the mutation operation to expand the local search range. The Pareto fast non-dominated sorting and congestion calculation are adopted to improve the solution quality and diversity, and the Pareto optimal solution is obtained step by step. The effectiveness of the improved genetic enhancement algorithm for the Pareto optimal solution of the multi-objective flow shop scheduling problem is verified by comparing the proposed algorithm with the genetic algorithm, NSGA-II algorithm, and Q-learning algorithm.

Key words: multi-objective flow shop scheduling; Q-learning; genetic algorithm; Pareto

随着生产规模的不断变大, 有效地安排各种资源成为了当前制造业信息化的主要任务. 多目标流水车

间调度问题 (multi-objective flow shop scheduling problem, MFSP) 是经典的生产调度问题^[1], 应用背景广泛, 在实

^① 收稿时间: 2023-08-23; 修改时间: 2023-09-26, 2023-10-20; 采用时间: 2023-10-25; csa 在线出版时间: 2023-12-27
CNKI 网络首发时间: 2023-12-29

际生产过程中,存在着多个作业同时进行且需要合理安排顺序以提高生产效率和资源利用率的需求,由于不同作业之间的关联性以及多个优化目标之间的冲突,如最小化生产时间、最大化利润、最小化设备闲置等,使得解决多目标车间调度 Pareto 最优问题变得非常复杂。

近年来,遗传算法 (genetic algorithm, GA), 粒子群算法 (particle swarm optimization, PSO) 等智能优化算法在求解车间调度问题领域取得很多研究成果。2009年,张国辉等^[2]提出了一种改进的遗传算法求解柔性作业车间调度问题。2010年, Gao 等^[3]提出了一种基于遗传算法的 GA-LS 算法求解采用部分作业序列集合的分布式置换流水车间调度问题,以最小化最大完工时间为优化目标,采用局部搜索方法寻找相邻解。2019年,宋存利^[4]针对混合流水车间调度问题,以最小化最大完工时间和最小化能耗为优化目标,建立混合整数线性规划模型,提出改进的快速非支配排序遗传算法。2023年,姜涛等^[5]提出了一种带有精英保留策略的遗传算法和模拟退火算法与文化算法框架结合的混合优化算法解决流水车间调度问题。谢美华等^[6]提出了一种自适应混合粒子群算法,结合 Q 学习设计了参数自适应策略以及引入粒子停滞判断方法求解置换流水车间调度问题。然而,现有智能优化算法求解车间调度 Pareto 最优解依然存在初始解质量不高、易陷入局部最优的问题。

近年来,强化学习在车间调度领域得到了很多关注。潘燕春等^[7]在 2007 年提出了一种采用强化学习优化个体的遗传强化学习算法,该算法应用于解决同序流水车间调度问题并得到了显著成果。2017年, Fonseca-Reyna 等^[8]提出一种使用 Q-learning 的强化学习算法用于解决流水车间调度问题。2019年,张东阳等^[9]应用强化学习算法,将组合优化问题转成序贯决策问题来解决置换流水车间调度问题。2021年,王凌等^[10]提出一种基于深度强化学习与迭代贪婪算法的框架,利用强化学习训练模型来获得更加优良的输出结果。王霄汉等^[11]详细地介绍了强化学习应用于车间调度问题的仿真技术和常用架构,指出了强化学习在解决车间调度问题有较低的时间响应和较优的模型泛化性。2022年,蔡静雯等^[12]设计了一种采用 Q-learning 算法的智能车间自适应调度算法来有效应对车间生产环境的变化,优化车间综合性能指标。2021年, He 等^[13]结合 Q-

learning 算法与 NHE 启发式算法,通过改变工件插入模式的方式提高算法效率。赫子淼^[14]研究了 Q-learning 算法在双目标 PFSP 中的应用根据车间的实际情况,以最大完工时间以及能量消耗作为优化目标。虽然强化学习在解决车间调度问题方面取得一些进展,但主要用于单目标问题,在求解多目标车间调度 Pareto 最优问题时存在求解效率和求解质量不高的问题。

求解多目标调度问题有两种常用的方法^[15]:一种方法是基于线性加权,将单个目标函数组合成一个复合函数,另一种是基于 Pareto 优先支配关系,将除一个目标外的其他目标存放在约束集中。与基于线性加权的优化方法对比, Pareto 优化方法不需要为每个优化目标分配权重^[16]。2006年, Rahimi-Vahed 等^[17]设计了一个利用理想点的新方法来指定最优粒子群在群中的位置向量,用于寻找多目标问题的局部 Pareto 最优前沿。2011年, Yang 等^[18]提出了一种基于遗传算法 (GA) 与基于禁忌搜索 (TS) 的 NPTSGA 算法,保持了 Pareto 最优解的无支配强化与近 Pareto 最优解在折中曲线上的平衡。2021年,张伟^[19]针对多目标流水车间生产调度问题,通过带有自适应参数调节改进的 NSGA-II 算法在生产调度解空间找寻 Pareto 解集,将数据进行加权处理可挑选出中意的调度方案。

针对多目标流水车间调度 Pareto 最优问题,本文设计了一种结合 Q-learning 算法的遗传强化算法 (genetic reinforcement algorithm with Q-learning, QGRA)。该方法利用 Q-learning 算法来获得初始种群,提高初始种群解的质量。在进化过程中综合使用多种交叉和变异算子,利用 Q-learning 算法得到的 Q 表对变异操作进行指导,提高算法跳出局部最优的能力。采用 Pareto 快速非支配排序和拥挤距离计算得到 Pareto 更优解,在进化过程中逐步获得 Pareto 最优解。

1 多目标流水车间调度问题

1.1 问题描述

MFSP 可描述为: n 个工件在 m 台机器上加工,每台机器加工一道工序,工件工序数即为 m ,且 n 个工件在 m 台机器上的加工顺序相同,本文的优化目标是 minimize 最大完工时间、最小化延误时间,确定每个工件在每台机器上的最优加工顺序约束条件如下。

- 1) 每台机器在同一时刻只能执行一道工序。
- 2) 每个加工的工件不能挂起或终止。

- 3) 每个工件在机器上的加工顺序相同。
4) 所有的运输时间和设置时间都忽略不计。

1.2 问题模型

MFSP 的参数以及决策变量定义如下。

J : 工件集合, $J = \{J_1, J_2, \dots, J_n\}$, n 为工件数。

M : 机器集合, $M = \{M_1, M_2, \dots, M_m\}$, m 为机器数。

$t_{i,j}$: 工件 J_i 在机器 M_j 上的加工时间。

$T_{i,j}$: 工件 J_i 在机器 M_j 上的完工时间。

C_i : 工件 J_i 的完工时间。

D_i : 工件 J_i 的交货期。

DDT : 工件紧急度。

$process_i$: 工件 J_i 的总处理时间。

$Tard_i$: 工件 J_i 的延误时间。

基于以上参数和变量定义, MFSP 的模型如下:

$$\begin{cases} T_{1,1} = t_{1,1} \\ T_{i,1} = T_{i-1,1} + t_{i,1}, & i \in [2, n] \\ T_{1,j} = T_{1,j-1} + t_{1,j}, & j \in [2, n] \\ T_{i,j} = \max\{T_{i-1,j}, T_{i,j-1}\} + t_{i,j}, & i \in [2, n], j \in [2, n] \end{cases} \quad (1)$$

$$C_i = T_{i,n} \quad (2)$$

$$C_{\max} = \max_{1 \leq i \leq n} C_i \quad (3)$$

$$process_i = \sum_{j=1}^m t_{i,j} \quad (4)$$

$$D_i = DDT \times process_i \quad (5)$$

$$Tard_i = \begin{cases} D_i - C_i, & D_i > C_i \\ 0, & D_i \leq C_i \end{cases} \quad (6)$$

$$Tard = \sum_{i=0}^n Tard_i \quad (7)$$

$$F_1 = \min\{C_{\max}\} \quad (8)$$

$$F_2 = \min\{Tard\} \quad (9)$$

其中, 式 (1) 计算每个工件在每台机器上的完工时间; 式 (2) 计算每个工件最大完工时间; 式 (3) 计算最大完工时间; 式 (4) 计算每个工件在所有机器上的总加工时间; 式 (5) 计算交货期^[20]; 式 (6) 计算每个工件的延误时间; 式 (7) 计算总的延误时间; 式 (8) 表示最小化最大完工时间目标; 式 (9) 表示最小化延误时间目标。

2 QGRA 算法设计

2.1 流水车间调度强化学习模型

强化学习算法应用到流水车间调度问题的关键是如何将问题映射到强化学习模型中. Q-learning 算法是经典的基于值函数的强化学习算法, 指在某个环境状态下, 采取某个动作作用于环境中, 环境根据所选择的动作反馈相应的奖励值. 本文采用 Q-learning 算法进行问题映射, 将当前加工的工件作为当前的环境状态, 选择下一工件的行为作为采取的动作, 被选择的工件为下一个新的环境状态, 由于按顺序加工的两个工件所产生的空闲时间属于负效益, 所以采用空闲时间的最大值减去相应工序的空闲时间作为其奖励值。

(1) 计算 R 表

R 表即环境根据动作反馈的奖励值, 本文将利用工件的空闲时间进行计算, 公式如下:

$$I_{i,k} = C_{i,k} - \sum_{j=1}^m t_{k,j} \quad (10)$$

$$R_{i,k} = \max(R) - I_{i,k} \quad (11)$$

其中, $I_{i,k}$ 表示从工件 i 到工件 k 的空闲时间, $C_{i,k}$ 表示从工件 i 到工件 k , 两个工件的最大加工完成时间, $t_{k,j}$ 表示工件 k 在机器 j 上的加工时间, $R_{i,k}$ 表示从工件 i 到工件 k 的奖励值。

(2) 计算 Q 表

Q 值是从状态 s 开始选择动作 a , 执行后获得的最大累积回报的折算值. Q-learning 算法的主要优势是使用了时间差分法 TD (融合了蒙特卡洛和动态规划) 能够进行离线学习, 使贝尔曼方程可以对马尔可夫过程求解最优策略。

$$Q(s, a)^{\text{new}} = (1 - \alpha)Q(s, a)^{\text{old}} + \alpha[R(s, a) + \gamma \max_{a'} [Q(s^{\text{new}}, a^{\text{new}})]] \quad (12)$$

$$Q(s, a)^{\text{old}} = R(s, a) + \gamma \max_{a'} Q(s^{\text{new}}, a^{\text{new}}) \quad (13)$$

其中, $\alpha \in (0, 1)$ 是学习率, $\gamma \in (0, 1)$ 是折扣率, s^{new} 是状态 s 在经过选择动作 a 后达到的新状态, 由式 (11) 可以看出, 当符合式 (13) 时, 迭代格式收敛, 也就是 Q-table 的更新完成。

具体训练过程如下。

首先, 随机选择一种状态, 将当前状态下所有可能的动作放到列表 r_action 中, 其次随机选择下一个状态并判断是否在 r_action 中, 若不在, 则重新选择下一

状态,若在,则将其在 r_action 中删除,采用式 (13) 计算 Q 值,当 r_action 的长度为 0 时,一次训练完成。

基于上述描述,本文建立了流水车间调度强化学习模型,如图 1 所示。

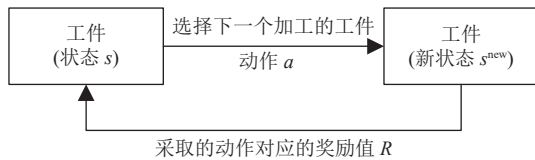


图 1 流水车间调度强化学习模型图

2.2 结合 Q 学习的种群初始化和变异操作

(1) 种群初始化过程

通过第 2.1 节计算得到的 Q 表,指导完成种群的初始化过程,获得初始种群。将可选的剩余工件放入临时列表 $temp$ 中,根据 Q 表的值,选择其中对应 Q 值最大的工件作为下一个加工工件并在临时列表 $temp$ 中移除,重复此步骤直到临时列表 $temp$ 为空,得到一个较优个体。本文分别将每个工件作为首个状态,由此可得到相应工件个数的较优个体,组成初始种群。例如,在一个有 9 个工件的加工任务中,工件 4 对应的 Q 表如表 1 所示,当初始加工任务是工件 4 时,选择临时列表 $temp$ 中所有剩余工件里面 Q 值最大的工件,工件 3 被选为下一个待加工工件,如图 2 所示。

表 1 工件 4 对应的 Q 表

Job4	1	2	3	4	5	6	7	8	9
Q值	26	19	96	-1	25	29	14	78	76

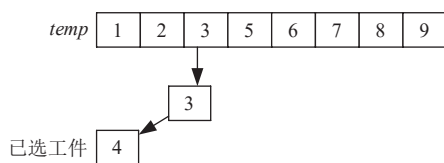


图 2 初始化工件选择示例图

(2) 变异过程

采用逆序操作、互换变异、插入变异 3 种变异方法进行变异操作,防止陷入局部最优状态。通过选择一种交叉算法来完成交叉操作。

变异操作是算法寻优和进化过程中重要的一步,以 RL 指导的插入变异操作为例,通过该方法能够提高算法的局部搜索能力,加快算法的收敛能力。随机选取一个变异位,执行变异操作后,删除 J_k 后续的子序列,根据 Q 表相应“状态-动作”的值,重新选择下一个要

加工的工件。GA 的子序列图如图 3 所示。

本文采用保优策略,判断子代个体相较于父代个体是否更优,若更优,则子代个体替换相应的父代个体,以此来将优良个体留存在种群当中,避免因算法的随机性导致的群体性能下降。

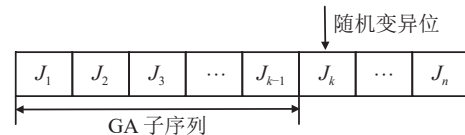


图 3 GA 子序列图

2.3 结合 Pareto 策略的选择操作

在求解多目标车间调度 Pareto 解的优化过程中,针对每一代进化群体,寻找出其当前最优个体,称一个进化群体的当前最优解为非支配解。估计当前种群某个个体周围个体的密度,维持种群中个体的多样性,拥挤距离越大越好,最后按照拥挤距离重新排序各层,进而排序种群。所有的非支配解的集合称为当前群体的非支配集,后续非支配集会不断的逼近最优解。

首先计算每个个体的适应度值,每个个体的适应度值由该个体的完工时间以及延误时间组成,然后根据快速排序的方法生成 Pareto 等级,快速非支配排序就是将解集分解为不同次序的 Pareto 前沿的过程,根据等级从低到高的顺序依次将整层种群放入新的种群中。根据每层个体拥挤距离从大到小排列,并以一定的概率选择较差解,否则选择较优解。

2.4 QGRA 算法流程

本文采用十进制的编码形式,编码后的染色体串为 (J_1, J_2, \dots, J_n) ,该编码方式组成的染色体本身即问题的一个解。其中,交叉部分采用线性次序交叉 (linear order crossover, LOX) 和部分交叉映射交叉 (partially mapping crossover, PMX) 方法进行交叉操作。

QGRA 求解 MFSP 的算法流程描述如算法 1。

算法 1. QGRA

输入: 种群规模、迭代次数 L 、工件紧急度 DDT ;
输出: 最优序列、Pareto 最优解。

- (1) 计算每个工件在每台机器上的加工时间,构建流水车间调度模型,初始化 R 表, Q 表;
- (2) 使用式 (10) 和式 (11) 计算相应的奖励值,更新 R 表;
- (3) 使用式 (13) 计算 Q 表,通过一定数量的迭代循环到达稳定值后,得到相应的折扣值。其中对折扣率 γ 的选取参考第 3.1 节仿真实验结果;
- (4) 把每个工件分别作为加工的第 1 个工件,由此产生相应数量的初始序列,并使用式 (1)-式 (3) 计算最大完工时间以及使用式 (4)-式 (7) 计算总的延误时间;

- (5) 计算适应度值;
- (6) 根据 Pareto 快速非支配排序以及拥挤度距离计算选择出两个个体, 并选择一种交叉方式对其进行交叉操作, 产生的新个体与旧个体进行比较, 采用保优策略, 将更优个体保留在种群中;
- (7) 根据 Pareto 快速非支配排序以及拥挤度距离计算选择出一个个体, 并选择一种变异方式对其进行变异操作, 产生的新个体与旧个体进行比较, 采用保优策略, 将更优个体保留在种群中;
- (8) 判断算法是否达到最大迭代次数, 如果达到最大迭代次数, 则终止算法, 输入最优值以及最优序列, 如果未达到则跳转到步骤 (5).

2.5 算法分析

对 QGRA 算法进行时间复杂度分析和空间复杂度分析, 对于强化学习部分, 计算 R 表的时间复杂度是 $O(n^2)$, 计算 Q 表的时间复杂度是 $O(G \times n)$, 由于 G 始终大于 n , 因此强化学习部分的时间复杂度为 $O(G \times n)$, 对于遗传算法部分, 由于线性交叉操作, 逆序变异操作以及计算最大完工时间的最大时间复杂度为 $O(n^2)$, 因此遗传算法部分的时间复杂度为 $O(G \times n^2)$, 整个算法的时间复杂度为 $O(G \times n^2)$, 由于整个算法控制选择的种群个体数为工件的数量, 因此整个算法的空间复杂度为 $T(n^2)$.

3 仿真实验与结果分析

本实验使用由 Taillard 等提供的标准 TA 类数据集, 选取了 12 种数据规模进行仿真实验. 相关参数设置如下: 迭代次数 $L=800$, 交叉概率 $P_c = 0.8$, 变异概率 $P_m = 0.1$. 由于当前数据集缺少对交货期的描述, 本文参考 Luo^[20]通过设置工件紧急度计算交货期的方法, 紧急度 DDT 的数值设定根据数据规模进行手工设定.

3.1 强化学习优化测试

强化学习的折扣率 γ , 通过实验证明, 不同规模的数据集, 当 γ 不同时, 得到的最优结果不同, 如表 2. 表 2 展示了通过 Q 学习产生的种群中最小的最大完工时间.

表 2 不同规模数据集在不同 γ 中运行测试结果

γ	20×5	20×10	50×5	50×10	100×5	100×10
0.1	1437	1935	3213	3785	6571	6812
0.2	1437	1962	3142	3785	6547	6828
0.3	1425	1988	3170	3756	6494	6701
0.4	1425	2040	3172	3760	6440	6696
0.5	1377	1972	3142	3724	6391	6760~
0.6	1437	1972	3143	3717	6319~	6737~
0.7	1437	1935	3129	3721	6195~	6660
0.8	1437	1879	3144	3566	6181~	6634
0.9	1377	1850	3105	3578	6134~	6591

注: ~表示数据值不稳定, 该值选取运行10次的平均值

由表 2 可以看出, 不同的数据规模对折扣率的要求不同, 当 $\gamma = 0.8$ 时, 50×10 规模数据得到的效果比其他折扣率较好, 当 $\gamma = 0.9$ 时, 其他规模数据仿真实验效果较好, 因此在接下来的算法优化测试中采用相应的最优折扣率.

本文选取最小规模 20×5 以及最大规模 100×10 进行初始化仿真实验, 分别采用随机方法和 Q-learning 方法产生初始种群, 计算其最大完工时间和总延误时间, 效果对比如图 4 所示.

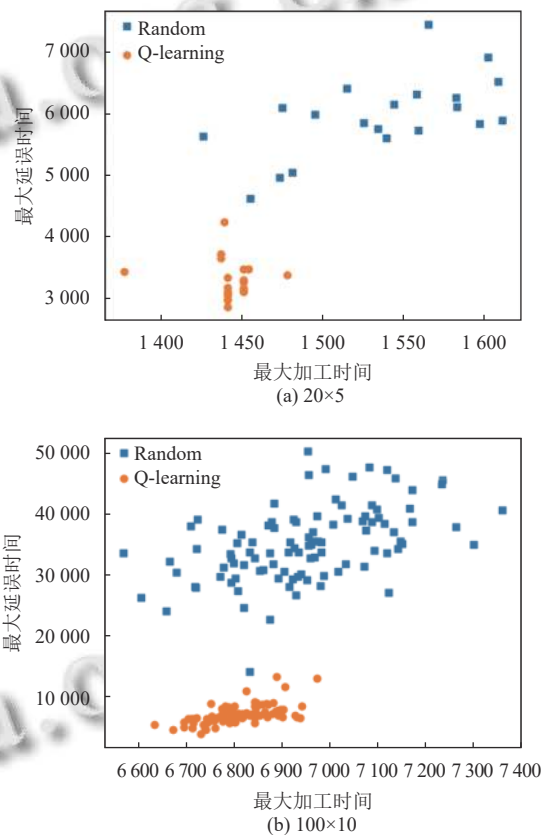


图 4 初始化效果对比图

由图 4 可以看出, 采用 Q-learning 方法得到的初始种群, 在最大完工时间和总的延误时间方面明显小于通过随机产生的初始种群, 同时解集比较集中于较优解, 初始种群能够得到较好的 Pareto 解. 因此, 通过 Q-learning 算法得到的初始种群质量有明显的提高.

3.2 实验结果与分析

为验证改进后的算法性能, 本文将 QGRA 算法与传统的遗传算法 (GA)、改进的自适应 NSGA-II 算法 (IANSGA)^[19]、改进的 Q-learning 算法 (IQL)^[14] 进行对

比实验. 4种算法的实验环境以及种群规模、迭代次数均相同, 各算法运行10次取最优解, 实验结果对比如

表3所示, 表3中最优解表示 $[F_1, F_2]$, 最优解值表示为 $best_result = 0.6 \times F_1 + 0.4 \times F_2$.

表3 各算法实验结果对比

规模	紧急度 (DDT)	GA		IANSQA		IQL		QGRA	
		最优解	最优解值	最优解	最优解值	最优解	最优解值	最优解	最优解值
20×5	[2, 3, 4]	[1377, 1544]	1460.5	[1377, 1765]	1571.0	[1377, 3330]	2518.2	[1377, 1533]	1439.4
20×10	[1.5, 2, 2.5]	[1867, 4542]	3204.5	[1814, 3385]	2442.4	[1654, 5559]	3216	[1717, 3134]	2425.5
50×5	[6, 7, 8]	[3156, 7431]	5293.5	[3329, 7388]	5358.5	[2976, 9602]	5626.4	[3310, 6826]	4364.8
50×10	[4.5, 5, 5.5]	[3526, 7522]	5124.4	[3517, 5356]	4436.5	[3331, 5424]	4168.2	[3465, 2641]	3135.4
100×5	[15, 17, 19]	[5891, 11115]	7980.6	[5915, 10707]	8311.0	[5598, 8979]	6950.4	[6060, 3532]	5048.8
100×10	[9, 10, 11]	[6827, 10577]	8327.0	[6575, 10465]	8520.0	[6282, 8549]	7188.8	[6639, 1081]	4418.2

由表3可以看出, 传统的GA和改进的自适应NSGA-II算法随着实验规模的增大, 最优解的质量越来越差, IQL算法在最大完工时间方面优势比较明显, 在总的延误时间方面优化效果较差, 因此最优解值较差, QGRA实验结果明显优于其他3种对比算法, 随着数量规模的增加, 优化效果更加明显.

图5列出了6种数据规模的运行效果对比图, 改进后的QGRA算法结合遗传算法和强化学习, 最大延误时间显著减小, 虽然最大完工时间结果略差于改进的IQL算法, 但综合后的最优解值是最好的. 随着数据规模的扩大, 改进后的QGRA算法优势逐渐明显, 因

此, QGRA能够有效求解多目标流水线调度Pareto最优问题.

4 结语

本文设计了一种基于Q-learning的遗传强化算法, 并用求解多目标流水线调度Pareto最优解. 该算法采用Q-learning算法获得高质量的初始解, 并利用Q表指导变异过程. 采用Pareto策略选择个体进行交叉变异操作, 提高算法跳出局部最优解的能力. 仿真实验结果说明, QGRA算法较好地解决了多目标流水线调度Pareto最优问题.

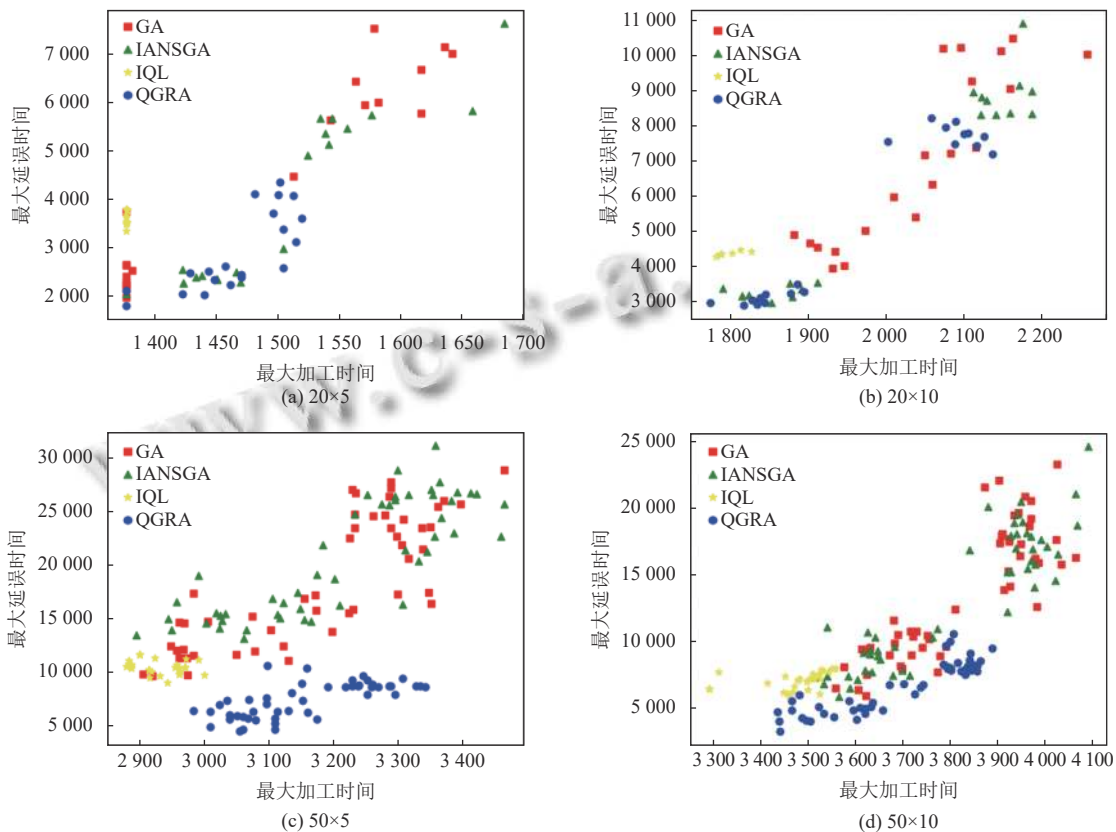


图5 不同规模的效果对比图

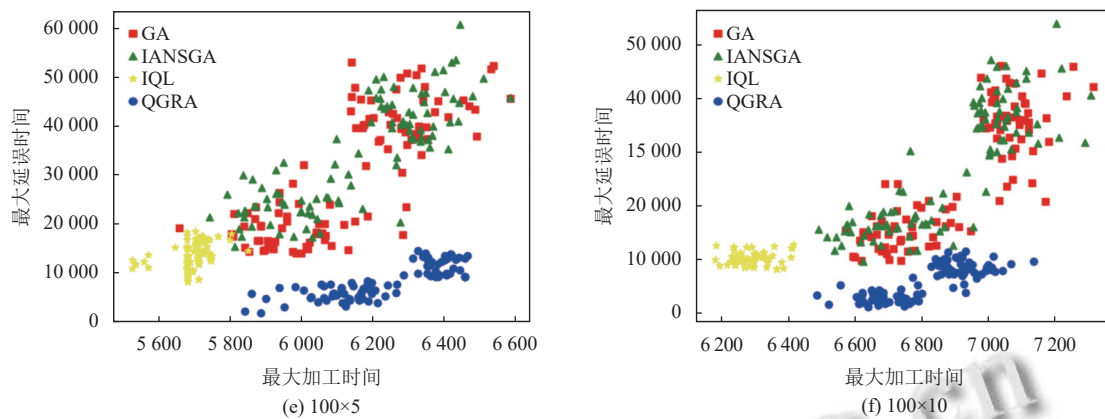


图5 不同规模的效果对比图(续)

参考文献

- 王凌. 车间调度及其遗传算法. 北京: 清华大学出版社, 2003.
- 张国辉, 高亮, 李培根, 等. 改进遗传算法求解柔性作业车间调度问题. 机械工程学报, 2009, 45(7): 145-151.
- Gao J, Chen R. A hybrid genetic algorithm for the distributed permutation flowshop scheduling problem. *International Journal of Computational Intelligence Systems*, 2012, 4(4): 497-508.
- 宋存利. 求解多目标混合流水车间调度的改进 NSGA-II. *计算机集成制造系统*, 2022, 28(6): 1777-1789. [doi: 10.13196/j.cims.2022.06.016]
- 姜涛, 周艳平. 混合文化优化算法及在车间调度中的应用. *计算机系统应用*, 2022, 31(12): 329-334. [doi: 10.15888/j.cnki.csa.008876]
- 谢美华, 李艳武, 葛棚丹. 自适应混合粒子群算法求解置换流水车间调度问题. *计算机应用研究*, 2023: 1-8. [doi: 10.19734/j.issn.1001-3695.2023.03.0105]
- 潘燕春, 周泓, 冯允成, 等. 同顺序 Flow-shop 问题的一种遗传强化学习算法. *系统工程理论与实践*, 2007, 27(9): 115-122. [doi: 10.3321/j.issn:1000-6788.2007.09.015]
- Fonseca-Reyna YC, Martínez-Jiménez Y, Nowé A. Q-learning algorithm performance for m-machine, N-jobs flow shop scheduling problems to minimize makespan. *Investigación Operacional*, 2018, 38(3): 281-290.
- 张东阳, 叶春明. 应用强化学习算法求解置换流水车间调度问题. *计算机系统应用*, 2019, 28(12): 195-199. [doi: 10.15888/j.cnki.csa.007196]
- 王凌, 潘子肖. 基于深度强化学习与迭代贪婪的流水车间调度优化. *控制与决策*, 2021, 36(11): 2609-2617. [doi: 10.13195/j.kzyjc.2020.0608]
- 王霄汉, 张霖, 任磊, 等. 基于强化学习的车间调度问题研究简述. *系统仿真学报*, 2021, 33(12): 2782-2791. [doi: 10.16182/j.issn1004731x.joss.21-FZ0774]
- 蔡静雯, 马玉敏, 黎声益, 等. 基于 Q 学习的智能车间自适应调度方法. *计算机集成制造系统*. <http://kns.cnki.net/kcms/detail/11.5946.TP.20220902.1128.002.html>. (2022-09-02)[2022-10-16].
- He ZM, Wang KL, Li HX, *et al.* Improved Q-learning algorithm for solving permutation flow shop scheduling problems. *IET Collaborative Intelligent Manufacturing*, 2022, 4(1): 35-44. [doi: 10.1049/cim2.12042]
- 赫子淼. 改进的 Q-Learning 算法求解置换流水车间调度问题研究 [硕士学位论文]. 聊城: 聊城大学, 2022.
- Konak A, Coit DW, Smith AE. Multi-objective optimization using genetic algorithms: A tutorial. *Reliability Engineering & System Safety*, 2006, 91(9): 992-1007.
- 罗哲. 面向多目标流水车间调度的混合遗传算法. *湖南科技学院学报*, 2017, 38(10): 71-74. [doi: 10.3969/j.issn.1673-2219.2017.10.025]
- Rahimi-Vahed AR, Mirghorbani SM. A multi-objective particle swarm for a flow shop scheduling problem. *Journal of Combinatorial Optimization*, 2007, 13(1): 79-102.
- Yang Y, Wu JF, Zhu XB, *et al.* A hybrid evolutionary algorithm for finding pareto optimal set in multi-objective optimization. *Proceedings of the 11th International Conference on Natural Computation*. Shanghai: IEEE, 2011. 1233-1236.
- 张伟. 改进的自适应 NSGA-II 求解多目标流水车间调度问题. *绵阳师范学院学报*, 2021, 40(5): 11-17. [doi: 10.16276/j.cnki.cn51-1670/g.2021.05.003]
- Luo S. Dynamic scheduling for flexible job shop with new job insertions by deep reinforcement learning. *Applied Soft Computing*, 2020, 91: 106208. [doi: 10.1016/j.asoc.2020.106208]

(校对责编: 牛欣悦)