



引用格式:程铭浩,唐浩,李宁,等.面向工艺匹配与规程切换的定制化生产多目标调度方法[J].海南大学学报(自然科学版中英文),DOI:10.65658/j.hndk.2026052103.

Citation:CHENG Minghao, TANG Hao, LI Ning, et al. A multi-objective scheduling method for customized production oriented to process matching and procedure switching[J]. Natural Science of Hainan University, DOI:10.65658/j.hndk.2026052103.

Open Access

面向工艺匹配与规程切换的定制化 生产多目标调度方法

程铭浩¹, 唐浩², 李宁²✉, 谢毅鹏², 王子峰², 徐博²

(1. 广东水利电力职业技术学院, 广东 广州 510925; 2. 海南大学 信息与通信工程学院, 海南 海口 570228)

摘要: 针对定制化生产中工艺路径差异、机器规程切换和资源分配耦合导致工序—机器匹配关系动态变化、多目标调度优化困难的问题, 本文建立面向工艺匹配与规程切换的定制化生产多目标调度模型, 并提出一种基于近端策略优化算法的直接调度决策方法。该方法将调度过程转化为马尔可夫决策过程, 利用工件加工信息和机器运行信息构建多矩阵状态表示, 以工序-机器组合作为动作输出, 并结合动作遮罩机制排除不可行动作, 通过目标导向奖励机制引导策略学习。实验结果表明, 所提方法在不同规模实例下的综合调度性能优于启发式规则和元启发式算法, 能够有效缩短最大完工时间、提高机器平均利用率并降低瓶颈机器负载, 在提升调度质量的同时保持较高的求解效率。

关键词: 定制化生产调度; 工艺匹配; 规程切换; 多目标优化; 近端策略优化

中图分类号: TP18 **文献标志码:** A

DOI: 10.65658/j.hndk.2026052103 **CSTR:** 32403.14.hndk.2026052103

A multi-objective scheduling method for customized production oriented to process matching and procedure switching

CHENG Minghao¹, TANG Hao², LI Ning²✉, XIE Yipeng², WANG Zifeng², XU Bo²

(1. Guangdong Polytechnic of Water Resources and Electric Engineering, Guangzhou 510925, China;

2. School of Information and Communication Engineering, Hainan University, Haikou 570228, China)

Abstract: In customized production, differences in process routes, machine procedure switching, and coupled resource allocation lead to dynamic operation-machine matching and make multi-objective scheduling difficult. To address this problem, this paper establishes a multi-objective scheduling model for customized production oriented to process matching and procedure switching, and proposes a direct scheduling decision-making method based on proximal policy optimization (PPO). The proposed method transforms the scheduling process into a Markov decision process. A multi-matrix state representation is constructed using job processing information and machine operation information, operation-machine combinations are used as action outputs, and an action mask mechanism is introduced to eliminate infeasible actions. An objective-oriented reward mechanism is further designed to guide policy learning. Experimental results show that the proposed method

Received: 2026-05-21; **Revised:** 2026-06-11; **Accepted:** 2026-06-16.

✉ **Corresponding author(s):** LI Ning, E-mail: lininggit@163.com.

achieves better comprehensive scheduling performance than heuristic rules and metaheuristic algorithms on instances of different scales. It can effectively shorten the makespan, improve average machine utilization, and reduce the bottleneck machine load while maintaining high solution efficiency.

Keywords: customized production scheduling; process matching; procedure switching; multi-objective optimization; proximal policy optimization.

0 引言

随着制造业由标准化、大批量生产向多品种、小批量和个性化方向发展,定制化生产逐渐成为制造企业提升市场响应能力和系统柔性的关键方式^[1]。与传统批量生产相比,定制化生产更强调以订单需求为牵引,根据产品属性、客户偏好和制造约束灵活组织加工过程^[2]。但是定制化生产中的工件结构、加工工序和制造规程往往存在明显差异,导致车间调度过程呈现出较强的异构性和复杂性。尤其在机器功能交叉、工序衔接紧密和资源并行运行的生产环境中,工件加工路径差异、机器规程切换以及资源竞争约束相互耦合,使调度系统不仅需要缩短最大完工时间,还需要兼顾机器利用率、瓶颈负载和调度方案可执行性等要求。

针对车间调度问题,现有研究已经形成了较为系统的求解方法,主要包括精确算法、启发式规则、元启发式算法和深度强化学习方法^[3]。精确算法建模严谨,能够在小规模问题中获得高质量解,适用于规模较小、约束结构相对清晰的调度场景,但随着问题规模和约束数量增加,其求解复杂度迅速上升,难以满足复杂生产场景中的快速决策要求。例如,Ham等^[4]针对带并行批处理机器的柔性作业车间调度问题建立了混合整数规划和约束规划模型。启发式规则具有实现简单、计算速度快等优势,在工程实践中应用较为广泛,适用于对响应速度要求较高的调度场景,但其性能通常依赖规则设计质量,不同规则对生产场景的适应性差异较大,且在多目标条件下容易出现对单一目标的偏向。例如,Panwalkar等^[5]系统总结了生产调度中的优先调度规则;Guo等^[6]进一步综述了基于超启发式的车间调度策略自动设计方法。元启发式算法通过种群进化、邻域搜索或随机扰动等机制增强全局搜索能力,能够获得较优调度方案,适用于复杂约束下的多目标离线优化问题,但往往需要较高的迭代计算开销,而且随着问题规模扩大,难以同时兼顾解的质量和响应效率。例如,An等^[7]研究了同时考虑新工件插入和机器预防性维护的多目标柔性作业车间重调度问题。

近年来,深度强化学习为复杂调度问题提供了新的求解思路^[8]。该类方法将调度过程建模为序列决策问题,使智能体能够通过与环境交互学习从生产状态到调度动作的映射关系,从而减少对固定规则的依赖^[9]。在深度强化学习调度方面,Luo^[10]针对新工件插入情形下的动态柔性作业车间调度问题提出深度强化学习方法;Wang等^[11]提出基于双注意力网络的强化学习方法,以刻画工序与机器之间的复杂关系;Lei等^[12]提出多动作深度强化学习框架,以支持柔性作业车间中工序与可加工机器之间的联合决策。上述研究表明,深度强化学习在状态感知、动作设计和动态调度方面具有较强潜力,并在柔性作业车间调度、动态作业调度和多目标排程等问题中取得了一定效果。但现有深度强化学习调度方法在定制化生产场景下仍存在不足:一方面,部分方法仍以启发式规则或复合规则作为动作输出,智能体本质上是在预设规则集合中进行选择,难以直接面向机器规程切换特征和工艺匹配约束完成工序与机器之间的细粒度匹配决策^[13]。另一方面,部分方法为降低输入维度,对生产状态进行较强压缩,导致工件加工进展、机器释放状态、候选工序差异和后续加工趋势等关键信息未能充分保留^[14]。在机器规程切换特征和工艺匹配约束共同作用下,若状态表达不够完整,策略网络便难以准确识别局部决策对规程切换代价、资源占用和整体排程结果的影响,从而削弱多目标协同优化能力。

综上所述,已有研究已从精确建模、规则调度、元启发式搜索和深度强化学习等方面推动了车间调度方法的发展,但对于具有机器规程切换特征和工艺匹配约束的定制化生产调度问题,现有方法在状态表达、切换代价建模、工序—机器直接决策和多目标协同优化方面仍存在不足^[15]。因此,本文建立面向工

艺匹配与规程切换的多目标调度模型,并提出一种基于近端策略优化算法(proximal policy optimization, PPO)的直接调度决策方法。该方法通过多矩阵状态表示、工序-机器组合动作空间、动作遮罩机制和目标导向奖励机制,实现最大完工时间、机器平均利用率和瓶颈机器负载的协同优化,并通过不同规模实例验证其有效性。

1 定制化生产调度问题描述

1.1 生产场景与调度问题描述 本文研究的定制化生产调度场景可描述为:车间需同时处理若干具有差异化工艺流程的工件,各工件包含数量不等的有序工序,且每道工序可由多台候选机器加工。由于机器加工能力、制造条件和工艺适配性不同,同一工序在不同机器上的加工时间存在差异。因此,调度过程不仅需要确定工序选择与机器分配方案,还需考虑工艺规程、机器状态与订单属性之间的匹配关系^[16]。

在该场景中,机器不仅承担工序加工任务,还具有相应的工艺规程状态。当工序所需规程与被分配机器的当前规程状态不一致时,机器需先完成规程切换并产生额外切换时间。因此,工件选择与机器分配并非独立决策,而是在工艺路径、机器状态、加工时间和切换代价共同作用下形成的耦合决策过程^[17]。

基于上述生产场景,本文研究的问题可描述为:在订单集合和机器资源已知的条件下,为各工件待加工工序选择加工机器并确定各机器加工顺序,使调度方案满足工序顺序约束、工序分配与机器互斥约束、工艺匹配与规程切换约束,并在最大完工时间、机器平均利用率和瓶颈机器负载3个目标之间进行权衡。与一般柔性作业车间调度主要关注工序顺序、候选机器选择、加工时间和机器互斥约束不同,本文所考虑的定制化生产调度还需刻画工序所需规程与机器当前规程状态之间的匹配关系。当二者不一致时,机器需进行规程切换并产生额外切换时间,使同一工序-机器组合在不同调度时刻可能具有不同实际代价。由此,当前工序-机器匹配决策不仅影响本工序的开始与完成时间,还会改变机器后续释放时间、规程状态和负载分布,从而增强调度过程的时序耦合性和多目标协同优化难度。

1.2 基本假设 为突出定制化生产中工序顺序、工序分配与机器互斥、工艺匹配与规程切换等约束对调度结果的影响,并保证模型构建的清晰性,作如下基本假设:

(1)所有待加工工件、工艺路线和机器资源在调度开始前均已知,机器初始时刻处于可用状态,且不考虑工件运输时间、机器故障、订单取消和新工件插入等动态事件。

(2)每个工件由若干有序工序组成,同一工件内部工序需按既定顺序加工,不同工件之间相互独立。

(3)每道工序可由其候选机器集合中的一台机器加工,且同一工序只能被加工一次。

(4)同一时刻每个工件只能在一台机器上加工,同一时刻每台机器只能加工一道工序。

(5)工序一旦开始加工,不允许中断或抢占。

(6)当工序所需规程与机器当前规程状态不一致时,机器需进行规程切换并产生相应切换时间。

1.3 多目标调度数学模型 为描述定制化生产调度模型,设车间共有 n 个工件和 m 台机器, J_i 表示第 i 个工件, M_k 表示第 k 台机器。工件 J_i 包含 n_i 道工序, O_{ij} 表示工件 i 的第 j 道工序, M_{ij} 表示该工序的候选机器集合。 t_{ijk} 、 S_{ijk} 和 E_{ijk} 分别表示工序 O_{ij} 在机器 M_k 上的加工时间、开始加工时间和完成加工时间。 $C_{\max}$ 、 U_k 和 L_k 分别表示最大完工时间、机器 M_k 的利用率和机器负载。 R_i 和 T_k 分别表示工件与机器的释放时间。 x_{ijk} 和 y_{ijk} 表示工序分配变量和规程切换变量,均为二值变量,表示 $(x_{ijk}, y_{ijk} \in 0, 1, \forall i, j, k)$ 。

(1)目标函数 在定制化生产调度过程中,工件工艺路径差异、机器加工能力差异和规程切换代价共同影响调度结果。为同时反应生产效率、资源利用和负载均衡要求,以最大完工时间、机器平均利用率和瓶颈机器负载构建多目标调度模型。同时,由于最大完工时间和瓶颈机器负载为最小化目标,而机器平均利用率为最大化目标,为统一优化方向,采用机器平均利用率倒数进行评价,目标函数表示为

$$\min F = \left(C_{\max}, \frac{1}{U_{ave}}, L_{\max} \right) = \left(\max_i C_i, \frac{1}{\frac{1}{m} \sum_{k=1}^m U_k}, \max_k L_k \right) \quad (1)$$

其中, C_i 表示工件 J_i 的完工时间, U_{ave} 表示机器平均利用率, L_{max} 表示瓶颈机器负载。目标函数(1)要求调度策略在缩短生产周期、提高机器利用率和抑制瓶颈机器负载集中之间进行协调优化。

(2)约束条件 为保证调度方案满足实际生产过程中的工序逻辑、资源占用和工艺规程要求,从工序顺序、工序分配与机器互斥、工艺匹配与规程切换、资源利用与负载计算 4 个方面构建约束条件。

首先,工序开始时间应该同时受到前序工序完成时间、机器释放时间和规程切换时间的约束,其时间递推关系可表示为

$$\begin{cases} S_{ijk} \geq \max(C_{i(j-1)}, T_k + y_{ijk}h_{ijk}), \\ E_{ijk} \geq S_{ijk} + t_{ijk}, \end{cases} \quad \forall i, j, k \quad (2)$$

其中, $C_{i,0} = R_i$, h_{ijk} 表示理想状态下机器 M_k 切换到工序 O_{ij} 所需工艺规程的时间。

其次,工序分配与机器互斥约束用于保证调度方案的资源分配可行性。一方面,每道工序必须且只能选择一台可加工机器完成,避免出现同一道工序被重复加工或未被加工的情况,即

$$\sum_{k \in \mathcal{M}_{ij}} x_{ijk} = 1, \forall i, j \quad (3)$$

另一方面,同一台机器在同一时刻只能加工一道工序。若 2 道不同工序 O_{ij} 和 $O_{i'j'}$ 被分配到同一台机器 M_k 上,则二者在加工时间上不得重叠,可表示为

$$\begin{cases} x_{ijk} = x_{i'j'k} = 1 \\ S_{ijk} + t_{ijk} \leq S_{i'j'k} + M(1 - z_{ijj'j'}) & ; \forall i, j, k, i', j', k'; i \neq i', j \neq j', k \neq k' \\ S_{i'j'k} + t_{i'j'k} \leq S_{ijk} + M \cdot z_{ijj'j'} \end{cases} \quad (4)$$

其中, M 为足够大的常数, $z_{ijj'j'}$ 为工序先后关系变量,用于表示同一机器上 2 道工序的加工顺序。式(3)保证工序分配的唯一性,式(4)则保证机器加工过程的互斥性。

工艺匹配与规程切换约束用于刻画定制化生产调度的关键特征。规程切换变量定义为

$$y_{ijk} = \begin{cases} 0, a_{ijk} = b_k \\ 1, a_{ijk} \neq b_k \end{cases}; \forall i, j, k \quad (5)$$

其中, a_{ijk} 表示工序 O_{ij} 在机器 M_k 上加工所需的规程编号, b_k 表示机器 M_k 当前规程状态。当二者不一致时,机器需先完成规程切换,并产生额外切换时间。

最后,机器利用率反映机器有效加工时间在总运行时间中的占比,其计算方式为

$$U_k = \frac{\sum_{i=1}^n \sum_{j=1}^{n_i} x_{ijk} t_{ijk}}{T_k}; \forall i, j, k \quad (6)$$

其中, T_k 表示机器 M_k 在当前调度方案中的最终释放时间,可用于表征该机器的总运行时间。机器负载由有效加工负载和规程切换负载共同构成,可表示为

$$L_k = \sum_{i=1}^n \sum_{j=1}^{n_i} (x_{ijk} t_{ijk} + y_{ijk} h_{ijk}); \forall i, j, k \quad (7)$$

其中 $x_{ijk} t_{ijk}$ 表示有效加工负载, $y_{ijk} h_{ijk}$ 表示由规程切换引起的额外负载。式(6)和式(7)为资源利用水平与机器负载评价提供计算依据。

2 面向工艺匹配与规程切换的多目标调度方法

2.1 调度过程的马尔可夫决策建模 基于 PPO 求解定制化生产调度问题,需要将调度过程转化为可交互学习的序列决策过程^[18]。在本文研究的定制化生产场景中,订单集合、工艺路径和机器资源在调度开始前已知,但每一次工序选择和机器分配都会改变工件加工状态、机器工作状态和规程状态。因此,当前

调度动作不仅影响局部工序的开始与完成时间,还会影响后续工序的可调度状态和整体多目标性能。基于上述特点,将定制化生产调度过程建模为马尔可夫决策过程:

$$M = (S, A, P, R, \gamma) \quad (8)$$

其中, S 为状态空间,用于描述当前车间生产状态; A 为动作空间,用于表示当前可选的工序-机器组合; P 为状态转移函数,用于描述执行调度动作后的车间状态更新过程; R 为奖励函数,用于评价调度动作对目标的影响; γ 为折扣因子,用于平衡当前奖励与未来累计收益。

在第 t 个决策时刻,智能体根据当前状态 $s_t \in S$ 选择动作 $a_t \in A$ 。该动作对应一次工序—机器匹配,即从当前候选工序中选择一道工序并分配至可加工机器。动作执行后,生产环境根据加工时间、机器释放时间和规程切换时间更新系统状态,得到下一状态 s_{t+1} ,并根据最大完工时间、机器利用率和加工时间变化等信息返回即时奖励 r_t 。该过程可表示为

$$s_{t+1} = P(s_t, a_t), r_t = R(s_t, a_t, s_{t+1}) \quad (9)$$

在上述建模框架下,调度策略的目标是学习从状态到动作的映射关系,使智能体在连续决策过程中获得更高累计回报。累计折扣回报可表示为

$$G_t = \sum_{\tau=t}^T \gamma^{\tau-t} r_\tau \quad (10)$$

其中, T 为一个完整调度过程的终止时刻。当所有工件的全部工序完成加工时,当前调度过程结束。通过最大化累计折扣回报,智能体可逐步学习不同车间状态下的工序选择和机器分配策略,从而实现最大完工时间、机器平均利用率和瓶颈机器负载等目标的综合优化。后续将分别从状态空间、动作空间、奖励机制和基于 PPO 的调度框架 4 个方面对该马尔可夫决策过程进行具体设计。

2.2 多矩阵状态空间设计 状态空间决定了智能体对生产环境的感知能力^[19]。在定制化生产场景中,工件工艺路径、机器加工能力和规程状态存在差异。若仅采用少量统计特征描述车间状态,容易造成候选工序加工差异、工件释放时间、机器释放时间和后续加工趋势等关键信息丢失。因此,采用多矩阵状态表示方法,将当前调度状态划分为工件信息矩阵和机器信息矩阵,以保留工件维度、机器维度和时间维度之间的结构关系^[20]。具体而言,状态空间由 4 个工件信息矩阵和 1 个机器信息矩阵组成,可以表示为

$$s_t = \{P_t, F_t, B_t, Z_t, Q_t\} \quad (11)$$

其中, P_t 为候选工序可选机器加工时间矩阵,用于描述各工件当前待加工工序在不同机器上的加工时间差异; F_t 为候选工序预估完成时间矩阵,用于刻画各工件在当前调度状态下的预估完成时间; B_t 为候选工序被调度后的预估剩余加工时间矩阵,用于反映当前候选工序被选择执行后,各工件剩余工序的预估加工压力; Z_t 为工件释放时间矩阵,用于记录每个工件能够被再次调度的最早时间; Q_t 为机器信息矩阵,机器信息矩阵用于描述各机器的释放时间,即各机器最早可用于加工新工序的时间。

其中,工件信息矩阵用于描述各工件当前候选工序的加工特征、完工趋势、剩余加工压力和释放边界。候选工序可选机器加工时间矩阵以工件为行、机器为列,元素表示当前候选工序在对应机器上的加工时间;若该工序不能在某台机器上加工,则用 -1 填充,若工件所有工序均已完成,则对应行用 0 填充。候选工序预估完成时间矩阵根据工件释放时间和剩余工序平均加工时间构造,用于表征不同工件的完工压力。预估剩余加工时间矩阵引入一步前瞻信息,用于反映当前候选工序被调度后各工件的后续加工压力。工件释放时间矩阵由工件释放时间列向量扩展得到,用于防止后续工序在前序工序完成前被调度。

机器信息矩阵由机器释放时间行向量扩展得到,能够反映机器空闲或繁忙状态,为判断候选工序能否在某台机器上尽早加工提供依据。当工序 O_{ij} 被分配至机器 M_k 后,机器释放时间根据工件释放时间、机器当前释放时间、加工时间和规程切换时间更新:

$$T_{k(\text{new})} = \begin{cases} T_{k(\text{old})} + t_{ijk} + y_{ijk} h_{ijk}, & T_{k(\text{old})} \geq R_i \\ R_i + t_{ijk} + \max(T_{k(\text{old})} + y_{ijk} h_{ijk} - R_i, 0), & T_{k(\text{old})} < R_i \end{cases} \quad (12)$$

当 $T_{k(\text{old})} \geq R_i$ 时, 说明机器释放时间不早于工件释放时间, 工件需要等待机器完成当前任务及可能发生的规程切换; 当 $T_{k(\text{old})} < R_i$ 时, 说明机器早于工件释放, 机器可在等待工件期间提前完成规程切换, 从而减少后续加工等待。通过上述多矩阵状态表示, 策略网络能够同时感知工件加工进展、候选工序加工差异、工件释放边界、机器释放状态和后续加工压力, 为后续工序-机器动作选择提供状态基础。

2.3 工序-机器动作空间与遮罩机制 动作空间设计直接影响策略搜索范围和调度性能^[21]。现有部分基于深度强化学习的调度方法将优先调度规则或复合规则作为动作输出, 智能体本质上是在预设规则集合中进行选择, 而非直接完成工序与机器之间的匹配决策。该方式虽然能够降低动作空间规模, 但策略性能容易受规则设计质量限制, 难以突破经验规则本身的搜索边界。

为提高策略的自主决策能力, 采用细粒度动作空间设计, 直接将当前候选工序与可选机器的组合作为智能体动作输出^[22]。在每个决策时刻, 每个未完成工件对应一道当前待加工工序 $J_i O$, 将该工序分配给某台机器 M_k 的动作记为 $A_{i,J_i O,k} = (J_i O, M_k)$ 。因此, 原始动作空间可定义为

$$A = \{(J_i O, M_k) | i = 1, 2, \dots, n; k = 1, 2, \dots, m\} \quad (13)$$

由于并非所有机器都能加工当前候选工序, 且已完成工件不应继续参与调度, 引入动作遮罩变量 $e_{i,J_i O,k}$ 对原始动作空间进行筛选:

$$e_{i,J_i O,k} = \begin{cases} 1, & J_i O \text{ 为当前候选工序且 } M_k \in \mathcal{M}_{i_j} \\ 0, & \text{其他} \end{cases} \quad (14)$$

当 $e_{i,J_i O,k} = 1$ 时, 表示工序 $J_i O$ 可以被分配至机器 M_k 加工; 当 $e_{i,J_i O,k} = 0$ 时, 表示该工序-机器组合不可行。具体而言, 遮罩机制仅保留各未完成工件的当前待加工工序, 并根据工序-机器可加工关系筛除不满足加工要求的机器。经过遮罩处理后, 策略网络仅在满足工序顺序约束和机器加工可行性约束的工序—机器组合中进行选择, 从而减少无效动作输出。由此, 经遮罩筛选后的可行动作空间 A_{valid} 可表示为

$$A_{\text{valid}} = \{(J_i O, M_k) \in A | e_{i,J_i O,k} = 1\} \quad (15)$$

当策略网络选择某一可行动作 $A_{i,J_i O,k} = (J_i O, M_k)$ 后, 将工件 J_i 的当前待加工工序 $J_i O$ 分配至机器 M_k , 并根据工序顺序、机器释放时间和工艺规程切换关系更新排程状态。该动作空间设计绕开了启发式规则中介, 使智能体能够直接学习不同生产状态下的工序选择与机器分配关系。对于工艺路径异构、机器规程状态频繁变化的定制化生产场景, 该动作空间设计有助于策略网络同时考虑加工时间差异、机器释放顺序、规程切换代价和负载分布, 从而提高调度决策的可行性与表达能力。

2.4 目标导向奖励机制 奖励机制用于评价智能体在当前状态下所选动作对调度目标的影响, 是引导策略网络更新的重要依据。结合已建立的多目标调度模型, 从最大完工时间、机器平均利用率和瓶颈机器负载 3 个方面设计目标导向奖励机制, 使智能体在交互过程中能够同时关注生产效率、资源利用和负载均衡。

在每个决策时刻, 智能体执行动作后, 环境状态由 s_t 转移至 s_{t+1} 。为反映动作对整体生产周期的影响, 将动作执行前后最大完工时间的变化作为基础奖励。为避免完工时间差值尺度过大导致训练波动, 将其除以所有工序中的最大加工时间进行归一化处理, 设 r_t^C 表示第 t 个决策时刻由最大完工时间变化得到的基础奖励, 则有

$$r_t^C = \frac{C_{\max(\text{old})} - C_{\max(\text{new})}}{\max(t_{ijk})} \quad (16)$$

其中, $C_{\max(\text{old})}$ 和 $C_{\max(\text{new})}$ 分别表示动作执行前后的最大完工时间。若动作执行后最大完工时间减小, 则基础奖励为正; 反之, 则基础奖励减小。

在基础奖励的基础上, 进一步引入资源利用和负载均衡反馈。若动作执行后机器平均利用率提高, 说明机器资源利用更加充分; 若瓶颈机器负载降低, 说明当前动作有助于缓解关键机器负载集中, 从而提升调度方案的均衡性。因此, 额外奖励项定义为

$$r_t^A = \begin{cases} 1, & U_{ave(new)} > 1.01 \cdot U_{ave(old)} \text{ or } L_{ave(new)} < 0.99 \cdot L_{ave(old)} \\ -1, & U_{ave(new)} < 0.99 \cdot U_{ave(old)} \text{ or } L_{ave(new)} > 1.01 L_{ave(old)} \\ 0, & \text{其他} \end{cases} \quad (17)$$

其中, $U_{ave(old)}$ 和 $U_{ave(new)}$ 分别表示动作执行前后的机器平均利用率, $L_{ave(old)}$ 和 $L_{ave(new)}$ 分别表示动作执行前后的瓶颈机器负载。式(17)中设置 1.01 和 0.99 作为阈值, 用于减少微小波动对奖励反馈的干扰。

综上, 当前决策动作对应的即时奖励可表示为

$$r_t = r_t^C + r_t^A \quad (18)$$

通过上述奖励机制, 智能体能够在每次动作执行后获得与调度目标相关的反馈信息。当动作有助于缩短最大完工时间、提高机器平均利用率或降低瓶颈机器负载时, 策略将获得正向强化; 当动作导致调度性能下降时, 策略将受到惩罚。目标导向奖励机制能够将多目标调度要求转化为连续交互过程中的即时反馈, 从而引导 PPO 学习更优的工序选择与机器分配策略。

2.5 基于 PPO 的调度框架 在完成状态空间、动作空间和奖励机制设计后, 采用基于 Actor-Critic 架构的 PPO 进行调度策略学习^[23]。在 PPO 框架下, 将多矩阵状态表示、工序-机器联合动作空间、动作遮罩机制和目标导向奖励机制统一起来, 使策略学习直接服务于工艺匹配、规程切换和多目标调度优化。

基于 PPO 的调度框架由特征提取网络、策略网络和价值网络组成。特征提取网络以多矩阵状态为输入, 将候选工序加工时间、工件释放时间、机器释放时间和剩余加工压力等信息编码为统一状态特征。策略网络根据该状态特征输出各工序-机器组合动作的选择概率, 并结合动作遮罩机制排除不可加工机器或已完成工件对应的无效动作。价值网络评估当前状态对后续调度目标的影响, 为策略更新提供参考。

训练阶段, 智能体根据当前状态选择可行的工序-机器组合动作, 环境根据工序加工时间、机器释放时间和规程切换时间更新工件加工进度、机器可用时间和规程状态, 并返回目标导向奖励。随后, PPO 根据交互数据更新策略, 使有利于提升调度质量的动作被强化。在策略更新过程中, PPO 通过裁剪目标函数限制新旧策略变化幅度, 以提高训练稳定性。

训练完成后, 所提方法不再需要像元启发式算法那样对每个实例进行长时间迭代搜索, 而是将当前生产状态输入训练好的策略网络, 直接输出工序-机器组合动作。随着状态不断更新, 策略网络逐步生成完整调度方案, 使多目标调度问题能够在兼顾最大完工时间、机器平均利用率和瓶颈机器负载的同时保持较高求解效率。整体调度框架如图 1 所示。

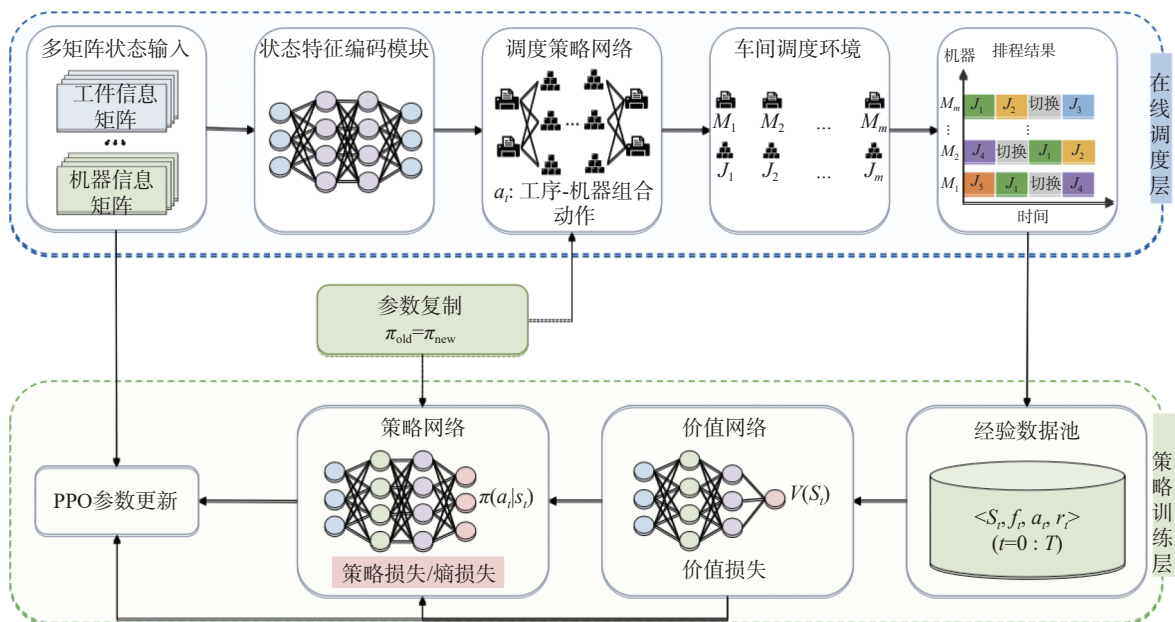


图 1 基于 PPO 的调度框架图

3 实验设计与结果分析

为验证所提方法在定制化生产多目标调度问题中的有效性,构建不同规模的生产实例,并与代表性启发式规则和元启发式算法进行对比。实验从调度方案可视化、多目标调度性能比较和求解效率分析3个方面展开,评价指标包括最大完工时间、机器平均利用率、瓶颈机器负载和计算时间。

3.1 实验设置与评价指标 为验证所提方法在不同规模定制化生产调度问题中的适用性,构建多组生产实例进行实验。实验实例由工件数量、机器数量、工序数量、工序加工时间、机器规程数和规程切换时间等参数生成,具体设置如表1所示。工件数量和机器数量按照由小到大的规模组合设定,以构造不同规模的调度实例,并考察算法在规模变化下的适用性;各工件的工序数量和工序加工时间在给定区间内随机生成,以反映定制化订单在工艺路线长度和加工负载上的差异;机器规程数和规程切换时间用于引入工艺匹配约束及规程切换代价,使生成实例能够体现本文所研究问题区别于一般柔性作业车间调度的动态耦合特征。为避免实例差异对算法比较造成影响,所有对比方法均在相同生产实例上进行测试。根据工件数量和机器数量,组合成 $n20 \times m5$ 、 $n30 \times m5$ 、 $n30 \times m8$ 、 $n50 \times m8$ 、 $n50 \times m15$ 、 $n80 \times m15$ 、 $n80 \times m20$ 和 $n100 \times m20$ 共8组代表性实例。

表1 生产车间参数

参数	数值	参数	设置
工件数量/个	[20, 30, 50, 80, 100]	工序加工时间/s	$U(20, 40)$
机器数量/台	[5, 8, 15, 20]	机器规程数/个	$U(3, 10)$
工序数量/道	$U(1, 20)$	工艺规程切换时间/s	$U(5, 15)$

注: $U(a, b)$ 表示参数在区间 $[a, b]$ 内按均匀分布随机生成;当参数为工序数量和机器规程数时,生成结果取整数;工序加工时间和工艺规程切换时间均以统一时间单位计。

PPO训练参数如表2所示。相关参数在参考PPO常用训练设置的基础上,结合本文状态空间、动作空间规模及预实验中的收敛情况确定。其中,网络层数和神经元数量根据多矩阵状态特征编码需求进行设置,以在保证模型表达能力的同时避免训练开销过大;学习率、折扣因子、裁剪因子及损失权重则结合预实验中累计奖励和主要调度指标的变化趋势进行选取,以保证训练过程的稳定性。为保证不同规模实验之间的可比性,在各组测试实例中保持PPO训练参数一致,未针对单一实例进行单独调整。

表2 PPO训练参数

参数	数值	参数	设置
特征网络/层	4	策略网络/层	3
特征网络输入层神经元/个	5	策略网络输入层神经元/个	128
特征网络隐藏层神经元/个	128	策略网络隐藏层神经元/个	64
特征网络输出层神经元/个	64	策略网络输出层神经元/个	1
价值网络/层	3	学习率	$1e-4$
价值网络输入层神经元/个	64	折扣因子	0.99
价值网络隐藏层神经元/个	64	每轮更新/次	1
价值网络输出层神经元/个	1	裁剪因子	0.2
训练总轮数/次	200	策略损失权重	3
价值损失权重	0.2	熵损失权重	0.05

注: 价值损失权重、学习率、折扣因子、裁剪因子、策略损失权重和熵损失权重均为无量纲训练参数。

选取最大完工时间、机器平均利用率、瓶颈机器负载和运行时间作为评价指标。其中,最大完工时间用于衡量整体生产周期;机器平均利用率用于衡量机器资源利用水平,实验比较时采用其倒数以保持优化方向一致;瓶颈机器负载用于衡量关键机器上的负载集中程度;运行时间用于衡量算法求解效率。上述指标能够从调度质量和求解效率2个方面评价所提方法的综合性能。

为了进一步验证所提方法的有效性,选取启发式规则和元启发式算法作为对比方法。启发式规则方面,选取 LOR(least operations remaining)、LWKR (least work remaining)和 SPT (shortest processing time)作为工件选择规则,分别从剩余工序数、剩余加工负载和当前工序加工时间角度确定调度优先级;选取 EST (earliest start time)和 LLM (least loaded machine)作为机器分配规则,分别从最早开始时间和机器负载均衡角度选择加工机器。由此形成 LOR_EST、LOR_LLM、LWKR_EST、LWKR_LLM、SPT_EST 和 SPT_LLM 6 种组合规则,以体现典型固定优先级规则在快速调度决策中的表现。元启发式算法方面,选取 MOPSO (multi-objective particle swarm optimization)、MOWOA (multi-objective whale optimization algorithm)和 NSGA-II(non-dominated sorting genetic algorithm II)作为对比对象,以体现典型群智能优化和进化多目标优化方法在复杂调度问题中的搜索能力。6 种启发式规则均按照“工件选择规则+机器分配规则”的方式实现,即在每个调度决策时刻先确定待调度工序,再从可加工机器集合中选择加工机器。3 种元启发式算法均采用由工序序列和机器选择序列构成的解表示,并通过相同解码过程计算最大完工时间、机器平均利用率倒数和瓶颈机器负载。为保证对比实验的公平性,各方法均在相同生产实例和相同评价指标下进行测试。3 种元启发式算法的种群规模和最大迭代次数均设置为 100,其他算法特有参数采用常用设置,并结合预实验中的运行稳定性进行确定。所有对比方法在每组实例下独立运行 20 次,并取各评价指标的平均值作为最终结果。

3.2 调度方案可视化分析 为直观分析所提方法生成调度方案的可执行性,选取由 20 个工件和 5 台机器构成的代表性实例进行甘特图可视化分析,结果如图 2 所示。图 2 中横轴表示加工时间,纵轴表示机器编号,不同色块表示被分配到相应机器上的加工工序,色块中的标注形式为“工件编号,工序编号”。灰色区段表示机器在加工不同工序前进行工艺规程切换所消耗的时间。

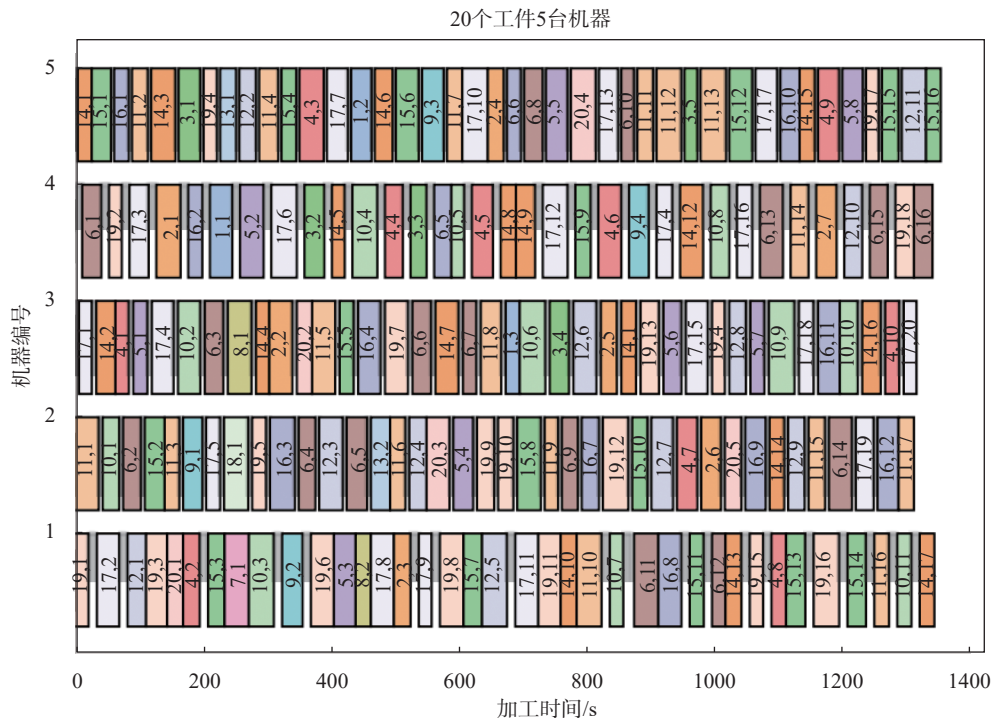


图 2 静态场景下的调度甘特图示例

由图 2 可以看出,所提方法生成的调度方案能够满足定制化生产调度的基本可行性要求。首先,同一工件内部各工序严格按照既定工艺顺序依次加工,未出现后续工序早于前序工序开始的情况,说明工序顺序约束得到满足;其次,同一台机器在同一时间段内仅执行一道工序,不同工序之间未发生时间重叠,说明机器互斥约束得到满足;最后,灰色规程切换区段表明调度方案能够根据工序所需规程与机器当前规程状态的匹配关系安排切换过程,从而体现了本文方法对工艺规程切换因素的考虑。

进一步观察可知,各机器上的加工任务整体衔接较为紧凑,较少出现长时间空闲区段。这说明基于 PPO 的调度策略能够结合当前工件释放时间、机器释放时间和规程切换状态进行工序—机器匹配决策,在保证工艺可行性的基础上提高机器时间资源的连续利用程度。由此可见,基于 PPO 的调度策略不仅能够生成满足约束条件的可执行调度方案,还能够在工序选择、机器分配和规程切换之间形成较为合理的协同安排。

3.3 多目标调度性能比较 为验证基于 PPO 的调度策略在定制化生产多目标调度问题中的有效性,将其分别与代表性启发式规则和典型元启发式算法进行对比。启发式规则由 3 种常用工件选择规则和 2 种机器分配规则组合,形成 LOR_EST、LOR_LLM、LWKR_EST、LWKR_LLM、SPT_EST 和 SPT_LLM 6 种复合动作规则,相关说明如表 3 所示。元启发式算法选取 MOPSO、MOWOA 和 NSGA-II。这些算法已被喻明让等^[24]、常大亮等^[25]、邹裕吉等^[26]、王思涵等^[27]和 Tang 等^[28]成功应用于各类调度问题,能够较好地体现对比算法的代表性与合理性。对比实验结果如表 4 和表 5 所示,评价指标包括最大完工时间($C_{\max}$)、机器平均利用率倒数($1/U_{ave}$)和瓶颈机器负载(M_{load})。

表 3 启发式规则

工件选择规则		机器分配规则	
LOR	选择剩余工序数最少的工件	EST	分配释放时间最早的机器
LWKR	选择剩余平均加工时间最短的工件	LLM	分配当前负载最小的机器
SPT	选择候选工序中平均加工时间最短的工件		

表 4 PPO 与启发式规则的对比实验

实例规模	指标	PPO	LOR_EST	LOR_LLM	LWKR_EST	LWKR_LLM	SPT_EST	SPT_LLM
$n20 \times m5$	$C_{\max}$	1.366e+03	3.957e+03	4.566e+03	3.885e+03	4.438e+03	2.958e+03	3.416e+03
	$1/U_{ave}$	1.234e+00	3.565e+00	4.074e+00	3.508e+00	3.965e+00	2.616e+00	3.018e+00
	M_{load}	1.366e+03	1.519e+03	1.485e+03	1.486e+03	1.502e+03	1.509e+03	1.506e+03
$n30 \times m5$	$C_{\max}$	2.210e+03	6.006e+03	6.699e+03	6.178e+03	6.774e+03	4.860e+03	5.588e+03
	$1/U_{ave}$	1.321e+00	3.604e+00	4.010e+00	3.654e+00	4.039e+00	2.914e+00	3.327e+00
	M_{load}	2.093e+03	2.156e+03	2.196e+03	2.300e+03	2.231e+03	2.238e+03	2.203e+03
$n30 \times m8$	$C_{\max}$	1.317e+03	4.711e+03	5.840e+03	4.824e+03	5.859e+03	3.579e+03	4.311e+03
	$1/U_{ave}$	1.246e+00	4.465e+00	5.541e+00	4.529e+00	5.564e+00	3.374e+00	4.058e+00
	M_{load}	1.317e+03	1.538e+03	1.486e+03	1.566e+03	1.483e+03	1.641e+03	1.512e+03
$n50 \times m8$	$C_{\max}$	2.325e+03	8.224e+03	1.0023e+04	8.181e+03	1.0056e+04	6.677e+03	8.107e+03
	$1/U_{ave}$	1.288e+00	4.511e+00	5.5380e+00	4.494e+00	5.5480e+00	3.642e+00	4.428e+00
	M_{load}	2.208e+03	2.654e+03	2.5660e+03	2.606e+03	2.5820e+03	2.612e+03	2.562e+03
$n50 \times m15$	$C_{\max}$	1.244e+03	5.553e+03	7.291e+03	5.592e+03	7.398e+03	3.764e+03	5.140e+03
	$1/U_{ave}$	1.293e+00	5.583e+00	7.567e+00	5.602e+00	7.482e+00	3.744e+00	5.097e+00
	M_{load}	1.241e+03	1.463e+03	1.372e+03	1.504e+03	1.372e+03	1.490e+03	1.352e+03
$n80 \times m15$	$C_{\max}$	1.866e+03	8.324e+03	1.1302e+04	8.218e+03	1.1209e+04	5.779e+03	7.912e+03
	$1/U_{ave}$	1.279e+00	5.505e+00	7.4960e+00	5.420e+00	7.4180e+00	3.837e+00	5.233e+00
	M_{load}	1.866e+03	2.240e+03	2.0310e+03	2.209e+03	2.0410e+03	2.206e+03	2.058e+03
$n80 \times m20$	$C_{\max}$	1.420e+03	6.737e+03	9.142e+03	6.828e+03	9.176e+03	4.636e+03	6.730e+03
	$1/U_{ave}$	1.279e+00	5.842e+00	7.938e+00	5.907e+00	7.960e+00	4.019e+00	5.801e+00
	M_{load}	1.419e+03	1.734e+03	1.621e+03	1.794e+03	1.620e+03	1.819e+03	1.637e+03
$n100 \times m20$	$C_{\max}$	2.076e+03	9.352e+03	1.2785e+04	9.382e+03	1.2885e+04	6.691e+03	9.432e+03
	$1/U_{ave}$	1.346e+00	6.098e+00	8.3340e+00	6.088e+00	8.3780e+00	4.293e+00	6.050e+00
	M_{load}	2.015e+03	2.298e+03	2.1380e+03	2.320e+03	2.1450e+03	2.299e+03	2.145e+03

表 5 PPO 与元启发式算法的对比实验

实例规模	指标	PPO	MOPSO	MOWOA	NSGA-II
$n20 \times m5$	$C_{\max}$	1.366e+03	1.630e+03	1.740e+03	1.534e+03
	$1/U_{\text{ave}}$	1.234e+00	1.549e+00	1.678e+00	1.381e+00
	M_{load}	1.366e+03	1.431e+03	1.453e+03	1.379e+03
$n30 \times m5$	$C_{\max}$	2.210e+03	2.413e+03	2.524e+03	2.277e+03
	$1/U_{\text{ave}}$	1.321e+00	1.523e+00	1.627e+00	1.367e+00
	M_{load}	2.093e+03	2.110e+03	2.229e+03	2.078e+03
$n30 \times m8$	$C_{\max}$	1.317e+03	1.714e+03	1.804e+03	1.656e+03
	$1/U_{\text{ave}}$	1.246e+00	1.709e+00	1.854e+00	1.547e+00
	M_{load}	1.317e+03	1.435e+03	1.487e+03	1.399e+03
$n50 \times m8$	$C_{\max}$	2.325e+03	2.753e+03	2.852e+03	2.675e+03
	$1/U_{\text{ave}}$	1.288e+00	1.609e+00	1.721e+00	1.465e+00
	M_{load}	2.208e+03	2.347e+03	2.454e+03	2.268e+03
$n50 \times m15$	$C_{\max}$	1.244e+03	1.875e+03	1.913e+03	1.786e+03
	$1/U_{\text{ave}}$	1.293e+00	2.004e+00	2.150e+00	1.779e+00
	M_{load}	1.241e+03	1.443e+03	1.454e+03	1.393e+03
$n80 \times m15$	$C_{\max}$	1.866e+03	2.616e+03	2.696e+03	2.533e+03
	$1/U_{\text{ave}}$	1.279e+00	1.849e+00	1.995e+00	1.654e+00
	M_{load}	1.866e+03	2.070e+03	2.091e+03	2.041e+03
$n80 \times m20$	$C_{\max}$	1.420e+03	2.158e+03	2.189e+03	2.078e+03
	$1/U_{\text{ave}}$	1.279e+00	2.005e+00	2.164e+00	1.779e+00
	M_{load}	1.419e+03	1.694e+03	1.640e+03	1.592e+03
$n100 \times m20$	$C_{\max}$	2.076e+03	2.825e+03	2.854e+03	2.731e+03
	$1/U_{\text{ave}}$	1.346e+00	1.978e+00	2.097e+00	1.748e+00
	M_{load}	2.015e+03	2.262e+03	2.276e+03	2.134e+03

由表 4 可知,相较于启发式规则,PPO 在多数实例中取得了更优的最大完工时间、机器平均利用率倒数和瓶颈机器负载。启发式规则通常依据固定优先级进行局部决策,难以同时兼顾工件释放时间、机器释放时间、加工时间和规程切换等因素。相比之下,PPO 基于多矩阵状态表示学习工序-机器组合决策策略,可以在不同规模实例下表现出更好的综合调度性能。

由表 5 可知,相较于 MOPSO、MOWOA 和 NSGA-II 等元启发式算法,PPO 在多数实例中同样具有较优表现,尤其在最大完工时间和机器平均利用率倒数方面优势较为明显。这说明 PPO 能够通过离线训练获得状态到动作的映射关系,在测试阶段直接生成调度决策,从而减少对反复迭代搜索的依赖。需要说明的是,在个别实例的瓶颈机器负载指标上,PPO 并非始终最优,表明其更侧重于在生产效率、资源利用和负载均衡之间取得综合折中。

综上所述,基于 PPO 的调度方法能够突破固定启发式规则的局部决策局限,并在多数实例中优于典型元启发式算法。实验结果表明,面向工艺匹配与规程切换关系设计状态空间、动作空间和奖励机制,有助于提升模型对复杂调度关系的表达能力,从而获得更优的多目标调度效果。

3.4 求解效率分析 为进一步分析所提方法的求解效率,对不同规模实例下各方法的计算时间进行比较,如图 3 所示。由图 3 可以看出,启发式规则的计算时间最短,这是因为它主要依据固定优先级规则完成工件选择和机器分配,不需要进行复杂搜索或策略更新。但结合多目标调度性能比较可知,启发式规则虽然具有较高计算效率,但调度质量容易受到固定规则适用范围的限制。

与启发式规则相比,元启发式算法需要通过种群迭代、个体更新和目标评价不断搜索较优解,因此计算时间明显增加,且随着工件数量和机器数量扩大,其计算开销呈上升趋势。相比之下,PPO在训练完成后,可通过策略网络前向推理直接输出工序-机器组合动作,不需要在测试阶段对每个实例进行长时间迭代搜索。因此,PPO的计算时间整体介于启发式规则与元启发式算法之间,并更接近启发式规则。

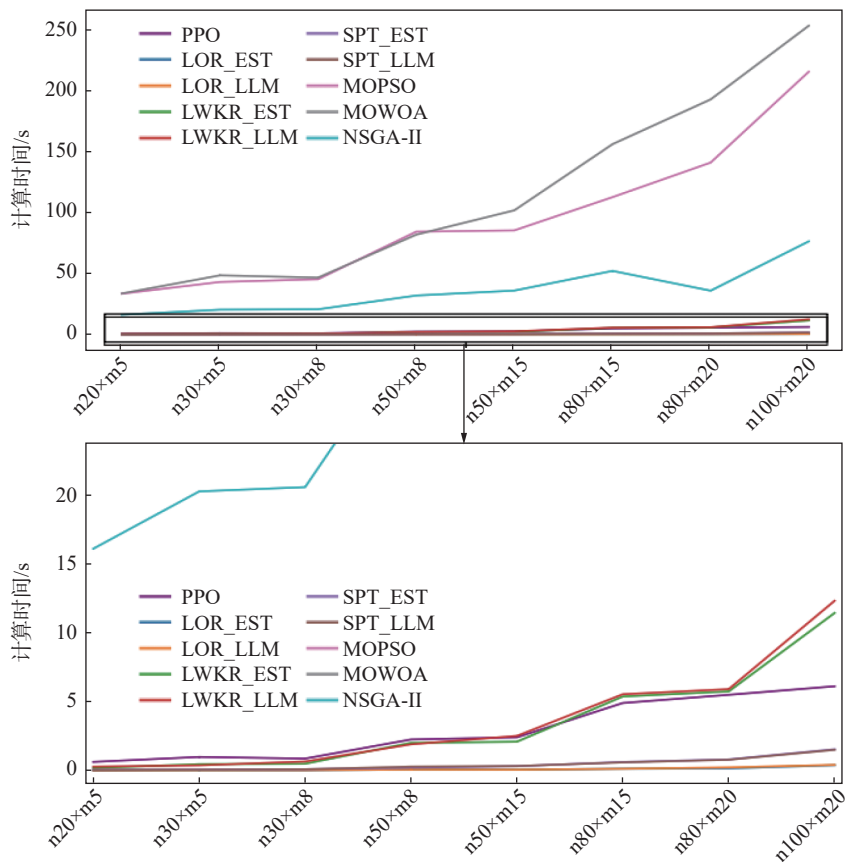


图3 计算时间对比图

综合图3、表3和表4可以看出,PPO在保持较高求解效率的同时,能够获得优于启发式规则和多数元启发式算法的多目标调度性能。这表明,将耗时的策略学习过程前置到离线训练阶段,并在在线调度阶段采用策略网络直接决策,有助于在调度质量和求解效率之间取得较好平衡。

4 结 论

针对定制化生产中工艺路径差异、机器规程切换频繁和资源分配关系复杂导致生产效率与资源均衡难以兼顾的问题,本文建立了面向工艺匹配与规程切换的多目标调度模型,并提出了一种基于近端策略优化算法的直接调度决策方法。该方法将多矩阵状态表示、工序-机器直接决策和目标导向奖励机制统一到强化学习框架中,使调度策略能够在生产状态感知、动作选择和多目标优化之间形成联动。实验结果表明,PPO能够在提升调度质量的同时保持较高的求解效率。由此可见,面向工艺匹配与规程切换关系进行建模,并将工序-机器匹配过程转化为可学习的直接决策过程,有助于突破固定启发式规则对调度策略的限制,提高模型对定制化生产复杂排程关系的表达能力。本文工作为定制化生产场景下兼顾生产效率、资源利用和负载均衡的智能排程提供了一种可行思路,后续研究可在此基础上进一步考虑动态订单到达、设备故障、工件运输以及多车间协同等因素,从而进一步增强模型对真实复杂制造环境的适应能力。

参考文献:

- [1] 季桓永, 苗仲桢, 王朋静, 等. 我国规模定制化生产典型模式与发展对策研究 [J]. *机械工程学报*, 2025, 61(10): 512–522.
JI H Y, MIAO Z Z, WANG P J, et al. Research on typical models and development strategies of mass customization in China[J]. *Journal of Mechanical Engineering*, 2025, 61(10): 512–522.
- [2] JOST P J, SÜSSER T. Company-customer interaction in mass customization[J]. *International Journal of Production Economics*, 2020, 220: 107454.
- [3] DAUZÈRE-PÉRÈS S, DING J W, SHEN L J, et al. The flexible job shop scheduling problem: a review[J]. *European Journal of Operational Research*, 2024, 314(2): 409–432.
- [4] HAM A M, CAKICI E. Flexible job shop scheduling problem with parallel batch processing machines: MIP and CP approaches[J]. *Computers & Industrial Engineering*, 2016, 102: 160–165.
- [5] PANWALKAR S S, ISKANDER W. A survey of scheduling rules[J]. *Operations Research*, 1977, 25(1): 45–61.
- [6] GUO H X, LIU J H, ZHUANG C B. Automatic design for shop scheduling strategies based on hyper-heuristics: a systematic review[J]. *Advanced Engineering Informatics*, 2022, 54: 101756.
- [7] AN Y J, CHEN X H, GAO K Z, et al. Multiobjective flexible job-shop rescheduling with new job insertion and machine preventive maintenance[J]. *IEEE Transactions on Cybernetics*, 2023, 53(5): 3101–3113.
- [8] 祝正宇, 郭具涛, 吕佑龙, 等. 面向柔性作业车间生产调度的深度强化学习方法 [J]. *中国机械工程*, 2024, 35(11): 2007–2014, 2034.
ZHU Z Y, GUO J T, LYU Y L, et al. Deep reinforcement learning method for flexible job shop scheduling[J]. *China Mechanical Engineering*, 2024, 35(11): 2007–2014, 2034.
- [9] YUAN E D, WANG L J, SONG S J, et al. Dynamic scheduling for multi-objective flexible job shop via deep reinforcement learning[J]. *Applied Soft Computing*, 2025, 171: 112787.
- [10] LUO S. Dynamic scheduling for flexible job shop with new job insertions by deep reinforcement learning[J]. *Applied Soft Computing*, 2020, 91: 106208.
- [11] WANG R Q, WANG G, SUN J, et al. Flexible job shop scheduling via dual attention network-based reinforcement learning[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, 35(3): 3091–3102.
- [12] LEI K, GUO P, ZHAO W C, et al. A multi-action deep reinforcement learning framework for flexible job-shop scheduling problem[J]. *Expert Systems with Applications*, 2022, 205: 117796.
- [13] DU Y, LI J Q, CHEN X L, et al. Knowledge-based reinforcement learning and estimation of distribution algorithm for flexible job shop scheduling problem[J]. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2023, 7(4): 1036–1050.
- [14] GUI Y, TANG D B, ZHU H H, et al. Dynamic scheduling for flexible job shop using a deep reinforcement learning approach[J]. *Computers & Industrial Engineering*, 2023, 180: 109255.
- [15] LI J H, LI X Y, GAO L, et al. A flexible job shop scheduling problem considering on-site machining fixtures: a case study from customized manufacturing enterprise[J]. *IEEE Transactions on Automation Science and Engineering*, 2025, 22: 8415–8426.
- [16] LI J H, LIU Q H, WANG C Y, et al. A disjunctive graph-based metaheuristic for flexible job-shop scheduling problems considering fixture shortages in customized manufacturing systems[J]. *Robotics and Computer-Integrated Manufacturing*, 2025, 95: 102981.
- [17] TENG G D. An improved genetic algorithm for dual-resource constrained flexible job shop scheduling problem with tool-switching dependent setup time[J]. *Expert Systems with Applications*, 2025, 281: 127496.
- [18] Schulman J, Wolski F, Dhariwal P, et al. Proximal policy optimization algorithms[EB/OL]. arXiv, (2017-07-20).
<https://arxiv.org/abs/1707.06347>. (查阅网上资料, 未找到本条文献引用日期信息, 请确认)
- [19] LUO S, ZHANG L X, FAN Y S. Dynamic multi-objective scheduling for flexible job shop by deep reinforcement learning[J].

- Computers & Industrial Engineering*, 2021, 159: 107489.
- [20] WU X Q, YAN X F. A spatial pyramid pooling-based deep reinforcement learning model for dynamic job-shop scheduling problem[J]. *Computers & Operations Research*, 2023, 160: 106401.
- [21] LIU C L, TSENG C J, HUANG T H, et al. Dynamic parallel machine scheduling with deep Q-network[J]. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2023, 53(11): 6792–6804.
- [22] SONG W, CHEN X Y, LI Q Q, et al. Flexible job-shop scheduling via graph neural network and deep reinforcement learning[J]. *IEEE Transactions on Industrial Informatics*, 2023, 19(2): 1600–1610.
- [23] YIMER H L, YANG P, QINGGE L. An improved actor-critic architecture with PPO for the traveling salesman problem[J]. *Expert Systems with Applications*, 2026, 298: 129723.
- [24] 喻明让, 陈云, 张志刚. 离散粒子群优化算法求解多目标柔性作业车间调度问题 [J]. *制造技术与机床*, 2019(1): 159–165.
- YU M R, CHEN Y, ZHANG Z G. Adiscrete version of particle swarm optimization for multi-objective flexible job-shop scheduling problems[J]. *Manufacturing Technology & Machine Tool*, 2019(1): 159–165.
- [25] 常大亮, 史海波, 孟树峰, 等. 有限缓冲区柔性流水车间调度优化问题求解 [J]. *制造业自动化*, 2024, 46(9): 159–169.
- CHANG D L, SHI H B, MENG S F, et al. Solving the optimal scheduling problem of flexible flow shop with limit buffer zone[J]. *Manufacturing Automation*, 2024, 46(9): 159–169.
- [26] 邹裕吉, 宋豫川, 王毅, 等. 基于离散型鲸鱼优化算法的 AGV 与机器集成调度方法 [J]. *重庆大学学报*, 2022, 45(6): 55–74.
- ZOU Y J, SONG Y C, WANG Y, et al. AGV and machine integrated scheduling method based on discrete whale optimization algorithm[J]. *Journal of Chongqing University*, 2022, 45(6): 55–74.
- [27] 王思涵, 黎阳, 李新宇. 基于鲸鱼群算法的柔性作业车间调度方法 [J]. *重庆大学学报*, 2020, 43(1): 1–11.
- WANG S H, LI Y, LI X Y. An improved whale swarm algorithm for flexible job-shop scheduling problem[J]. *Journal of Chongqing University*, 2020, 43(1): 1–11.
- [28] TANG H, WANG Z F, ZHANG Y H, et al. A sustainable computing-based multi-objective feeding scheduling model for low-carbon factory aquaculture using enhanced NSGA-II in agricultural consumer electronics[J]. *IEEE Transactions on Consumer Electronics*, 2025, 71(2): 6989–7001.

作者贡献声明

本声明旨在明确各作者对本论文的具体贡献, 确保学术成果署名的规范性和公平性, 各作者贡献如下: 程铭浩负责研究构思与总体方案设计, 完成调度模型构建、算法实现、实验分析、论文初稿撰写及全文修改工作。唐浩负责研究方向把握、理论框架指导和经费支持, 对论文结构设计、方法合理性和研究结论提出重要修改意见。李宁负责研究框架完善、技术路线论证、实验结果解释、论文修改、投稿沟通及最终稿件审核。谢毅鹏参与文献整理、实验数据处理、结果统计分析和论文部分内容修改。王子峰参与仿真实验运行、对比实验整理、评价指标计算和图表制作。徐博参与论文内容审阅、逻辑结构优化、语言润色和关键修改意见提出。

利益冲突声明

作者声明, 其不存在已知的可能影响本论文所报告工作的竞争性经济利益或个人关系。

AI 使用声明

本文未使用人工智能(AI)协助撰写。

伦理声明

不适用: 本文不涉及人类或动物的生物学研究。

电子补充材料

无

数据可用性

支持论文结论所需的所有数据均已包含在稿件中。如需获取与本文相关的其他数据,可向通信作者提出申请。

致谢/Acknowledgements

本文得到国家自然科学基金项目(62563008、62303142)和海南省重点研发项目(ZDYF2025GXJS183)支持。

This work was supported by the National Natural Science Foundation of China (62563008, 62303142) and the Key Research and Development Program of Hainan Province (ZDYF2025GXJS183).

作者简介

程铭浩(1998—),男,广东高州人,讲师,研究方向:具身智能机器人、视觉伺服等。E-mail: minghao_cheng@163.com。

唐 浩(1992—),男,海南东方人,博士,副教授,博士生导师,研究方向:智能装备管控等。E-mail: melineth@hainanu.edu.cn。

李 宁(1998—),男,黑龙江鸡西人,海南大学信息与通信工程专业2026级博士研究生。E-mail: lininggit@163.com。

谢毅鹏(2006—),男,福建莆田人,海南大学电子信息工程专业2025级本科生。E-mail: 13718886183@163.com。

王子峰(2000—),男,广东梅州人,海南大学信息与通信工程专业2023级硕士研究生。E-mail: wangzifeng@hainanu.edu.cn。

徐 博(1994—),女,吉林长春人,博士,副教授,博士生导师,研究方向:动态信号检测和处理等。E-mail: 996458@hainanu.edu.cn。

(责任编辑:高 喆)



This is an open access article under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0, <https://creativecommons.org/licenses/by/4.0/>).

© The Author(s) 2026. Published by Journal Editorial Department of Hainan University.

CHENG Minghao et al. A multi-objective scheduling method for customized production oriented to process matching and ...

Natural Science of Hainan University (ISSN 1004-1729)

<https://hndk.hainanu.edu.cn>