

文章编号:1006-2467(2022)09-1262-14

DOI: 10.16183/j.cnki.jsjtu.2021.215

# 面向柔性作业车间动态调度的双系统强化学习方法

刘亚辉<sup>1</sup>, 申兴旺<sup>1</sup>, 顾星海<sup>1</sup>, 彭涛<sup>2</sup>, 鲍劲松<sup>1</sup>, 张丹<sup>1</sup>

(1. 东华大学 机械工程学院, 上海 201620; 2. 浙江大学 机械工程学院, 杭州 310027)

**摘要:** 航天结构件生产过程中批产任务与研发任务并存, 个性化小批量研发生产任务导致紧急插单现象频发. 为了保障任务如期完成, 解决柔性作业车间面临的动态调度问题, 以最小化设备平均负载和最小化总完工时间为优化目标, 提出了感知-认知双系统驱动的双环深度 Q 网络方法. 感知系统基于知识图谱实现对车间知识的表示并生成多维信息矩阵; 认知系统将调度过程分别抽象为资源配置智能体和工序排序智能体两个阶段, 分别对应两个优化目标, 设计了车间状态矩阵对问题和约束进行描述, 调度决策中分步骤引入动作指令; 最后分别设计奖励函数实现资源配置决策和工序排序决策的评价. 经某动力所航天壳体加工的实例验证和算法对比分析, 验证了所提方法的优越性.

**关键词:** 感知-认知双系统; 双环深度 Q 网络; 动态调度; 知识图谱; 多智能体

**中图分类号:** TP 301 **文献标志码:** A

## A Dual-System Reinforcement Learning Method for Flexible Job Shop Dynamic Scheduling

LIU Yahui<sup>1</sup>, SHEN Xingwang<sup>1</sup>, GU Xinghai<sup>1</sup>, PENG Tao<sup>2</sup>, BAO Jinsong<sup>1</sup>, ZHANG Dan<sup>1</sup>

(1. School of Mechanical Engineering, Donghua University, Shanghai 201620, China;

2. School of Mechanical Engineering, Zhejiang University, Hangzhou 310027, China)

**Abstract:** In the production process of aerospace structural parts, there coexist batch production tasks and research and development (R&D) tasks. Personalized small-batch R&D and production tasks lead to frequent emergency insertion orders. In order to ensure that the task is completed on schedule and to solve the flexible job shop dynamic scheduling problem, this paper takes minimization of equipment average load and total completion time as optimization goals, and proposes a dual-loop deep Q network (DL-DQN) method driven by a perception-cognition dual system. Based on the knowledge graph, the perception system realizes the representation of workshop knowledge and the generation of multi-dimensional information matrix. The cognitive system abstracts the scheduling process into two stages: resource allocation agent and process sequencing agent, corresponding to two optimization goals respectively. The workshop status matrix is designed to describe the problems and constraints. In scheduling decision, action instructions are introduced step by step. Finally, the reward function is designed to realize the evaluation of resource allocation decision and process sequence decision. Application of the proposed method in the aerospace shell processing of an aerospace institute and comparative analysis of different algorithms verify

收稿日期: 2021-06-22

基金项目: 国家重点研发计划(2019YFB1706300)资助项目

作者简介: 刘亚辉(1997-), 女, 河南省许昌市人, 硕士生, 从事认知制造、知识图谱、智能调度研究.

通信作者: 鲍劲松, 男, 教授, 博士生导师, 电话(Tel.): 021-67792567; E-mail: bao@dhu.edu.cn.

the superiority of the proposed method.

**Key words:** perception-cognition dual system; dual-loop deep Q network (DL-DQN); dynamic scheduling; knowledge graph; multi-agent

航天结构件的加工过程具有多品种、小批量、工艺复杂多变的特征,每种工件由多道具有先后顺序约束的工序完成,是典型的作业车间调度问题。由于航天结构件紧急插单任务的扰动影响批量生产任务的生产进程,降低生产效率,所以车间调度方案需要动态变更。在传统作业车间调度的基础上,柔性作业车间调度可以面向更加动态、复杂的车间环境。根据柔性作业车间的特点,可以将该类调度问题拆分为资源配置和工序排序两个子问题,分别利用调度系统中存在的多类型知识,赋予系统自适应响应能力。

柔性作业车间调度需要利用实时性状态数据来驱动,但目前系统对于车间资源中离散程度高、关联性差的数据无法很好地利用。在处理大规模复杂数据方面,知识图谱技术<sup>[1-2]</sup>能够表达更丰富的语义,对数据进行语义信息集成与表示。认知制造<sup>[3]</sup>作为下一代智能制造<sup>[4]</sup>,利用感知-认知系统实现信息-物理系统的交互。知识图谱技术与认知制造的结合应用增强了车间动态调度策略优化与知识之间的联通性,一定程度上提高了决策方案的科学性。如何利用知识图谱提供的调度知识实现动态调度决策优化是值得研究的课题。

近年来,已有众多学者针对动态柔性作业车间调度问题<sup>[5]</sup>进行研究,主要在鲁棒式方法、重调度法、智能算法等方面展开。鲁棒式方法<sup>[6-7]</sup>对生产扰动事件的响应速度慢,历史信息无法得到重用;重调度方法<sup>[8]</sup>可以对扰动事件快速响应,但无法很好利用工艺信息之间的关系;相比之下智能算法<sup>[9-10]</sup>更适用于处理作业车间的动态调度问题。通过智能算法完成自适应车间调度问题的方法有优先级规则方法<sup>[11]</sup>、启发式算法<sup>[12]</sup>、遗传算法(Genetic Algorithm, GA)<sup>[13]</sup>、进化算法<sup>[14]</sup>、模拟退火<sup>[15]</sup>、禁忌搜索方法<sup>[16]</sup>等。在考虑插单任务的动态调度研究中,Kundakci等<sup>[17]</sup>针对动态作业车间调度问题中的插单任务、机器故障等,引入有效的混合遗传算法求解此类问题的最大完工时间;Shen等<sup>[18]</sup>针对柔性作业车间调度的新工件插入和机床故障等问题,提出了基于多目标进化算法的主动-被动调度方法;Wang等<sup>[19]</sup>提出了一种改进的粒子群算法用于工件随机到达的动态柔性作业车间调度问题的求解。为了实现多约束下调度问题的多目标优化,一些研究

学者将调度过程进行分步骤处理。张洁等<sup>[20]</sup>提出针对调度过程的任务分派和任务排序两阶段,构建具有继承关系的两阶段蚁群并行搜索算法;周亚勤等<sup>[21]</sup>提出一种嵌套式蚁群-遗传混合算法解决关键设备和人员双资源约束下的航天结构件柔性作业车间调度问题;汪浩祥等<sup>[22]</sup>利用强化学习的实时性和知识化制造系统的自适应性,将航空发动机装配调度分为分派机器和工序调度两个阶段。现有的调度规则大多无法进行自适应调整以应对动态场景,但随着机器学习的快速发展,可通过学习生成复杂调度策略的强化学习方法已经产生,能赋予调度规则较强的自适应和学习能力。进一步,深度Q网络(Deep Q Network, DQN)结合了具有感知能力的深度学习和决策能力的强化学习,融合了深度神经网络能够利用历史数据在线学习和强化学习从而依据状态灵活选取决策行为的优点,更适用于柔性作业车间调度的决策过程<sup>[23]</sup>。Wang等<sup>[24]</sup>在多智能体强化学习环境中运用DQN指导基础设施云上多工作流的调度,分别将最大完工时间和成本两个优化目标抽象成两个智能体。Luo<sup>[25]</sup>研究了用DQN解决以总延期最小为目标的新任务插入下的动态柔性作业车间调度问题。He等<sup>[26]</sup>针对纺织过程优化问题转化为随机博弈的新框架,在DQN网络算法中将多优化目标抽象成多智能体。林时敬等<sup>[27]</sup>针对炼钢车间天车动态调度事件,也运用DQN算法将多天车抽象成多智能体协同天车操作动作序列。但现在基于DQN的车间调度方法中,所需调度策略的优化过程与调度相关数据的关联度不够,因此考虑在调度算法中利用系统提供的语义知识优化车间动态调度能力。

为此,本文针对柔性作业车间调度问题,以最小化设备平均负载和最小化总完工时间为目标,在调度过程的资源配置和工序排序阶段,分别设置强化学习智能体,提出双系统强化学习框架。其中,基于感知系统获取知识图谱的知识,分别将资源配置和工序排序所需要的参数集成表示为多维信息矩阵的形式,为认知调度系统提供车间状态参数;认知调度过程中对两个智能体共同构建具有继承关系的双环深度Q网络(Dual-Loop Deep Q Network, DL-DQN)方法。

1 柔性作业车间调度模型

柔性作业生产车间调度问题可以描述为:总共有  $K$  个任务订单,其中包括批产任务和插单任务,每个任务包含  $n$  个待加工工件,即  $J = \{J_1, J_2, \cdots, J_i\} (i = 1, 2, \cdots, n)$ ,所有任务动态到达需要在  $L$  种设备组  $G = \{G_1, G_2, \cdots, G_l\} (l = 1, 2, \cdots, L)$  上加工,每个设备组包含一定数量  $T$  的设备  $M = \{M_1, M_2, \cdots, M_t\} (t = 1, 2, \cdots, T)$  执行对应的工序,不同的任务包含不同的加工工件数目、对应不同的加工工序,需要在设备组上进行工序排序并加工.对插单任务进行资源配置,并根据资源配置结果对机加工工序进行优化排序,使得设备总负载和任务总完工时间最小.在问题描述中涉及到的参数如表 1 所示.

表 1 符号与变量  
Tab. 1 Symbols and variables

符号	符号描述
$J$	工件集合
$G$	设备组集合
$M$	设备集合
$P$	人员集合
$S$	物料集合
$o_{k,i,j}$	第 $k$ 个任务中工件 $J_i$ 的第 $j$ 道工序
$j$	工序索引 $j=1, 2, \cdots, m$
$R_{k,i,j}$	第 $k$ 个任务中工件 $J_i$ 的第 $j$ 道工序配置资源, $R_{k,i,j} = \{M_{k,i,j}, P_{k,i,j}, S_{k,i,j}\}$
$M_{k,i,j}$	工序 $o_{k,i,j}$ 的配置设备
$P_{k,i,j}$	工序 $o_{k,i,j}$ 的配置操作人员
$S_{k,i,j}$	工序 $o_{k,i,j}$ 的配置物料
$S_{k,i}^T$	第 $k$ 个任务中工件 $J_i$ 的到达时间
$s_{o_{k,i,j}}^t$	工序 $o_{k,i,j}$ 的开始时间
$e_{o_{k,i,j}}^t$	工序 $o_{k,i,j}$ 的结束时间
$r_{o_{k,i,j}}^t$	工序 $o_{k,i,j}$ 与下道工序的准备时间
$B_{k,i}^T$	第 $k$ 个任务中工件 $J_i$ 的总加工时间
$W_{l,t}(W_{k,i})$	设备组 $G_l$ 中设备 $M_t$ 的加工负载(以设备为目标计算得到 $W_{l,t}$ ,以工序为单位计算得到 $W_{k,i}$ )
$L_{l,t}^W$	设备组 $G_l$ 中设备 $M_t$ 的最大加工负载
$B_{M_{k,i,j}}^{J_{k,i}}$	0-1 决策变量,取 1 时表示第 $k$ 个任务中工件 $J_i$ 在设备 $M_{i,j}$ 上加工
$D^P$	交交期

柔性作业车间调度模型需要满足如下基本假设:①批产任务中工件一定的情况下,插单任务的工件动态到达;②同一工件工序之间有优先级,需在前

道工序加工完成后进行后道工序的加工;③工件某一道工序在加工过程中不能被中断;④ 工件在加工过程中的准备时间、工件转换时间等一同计入加工时间;⑤同一设备在某一时刻只能加工一道工序;⑥ 同一设备前一工件未完成,后续安排工件需等待;⑦不存在撤单、设备故障等特殊情况.

建立如下数学模型:

$$\min f_1 = \min \left( \frac{1}{N_M} \sum_{l=1}^L \sum_{t=1}^T W_{l,t} \right) \tag{1}$$
$$\forall l \in [1, 2, \cdots, L], t \in [1, 2, \cdots, T]$$

$$\min f_2 = \min \left( \sum_{k=1}^K \sum_{i=1}^n B_{k,i}^T \right) \tag{2}$$
$$\forall i \in [1, 2, \cdots, n], k \in [1, 2, \cdots, K]$$

式中:

$$W_{l,t} = W_{k,i} = \sum_{j=1}^m B_{M_{k,i,j}}^{J_{k,i}} (e_{o_{k,i,j}}^t - s_{o_{k,i,j}}^t)$$
$$B_{k,i}^T = S_{k,i}^T + \sum_{j=1}^m B_{M_{k,i,j}}^{J_{k,i}} (e_{o_{k,i,j}}^t - s_{o_{k,i,j}}^t + r_{o_{k,i,j}}^t)$$

并满足如下约束:

$$1 \leq \sum_{j=1}^m \sum_{i=1}^n B_{M_{k,i,j}}^{J_{k,i}} \leq N_{\text{num}} \tag{3}$$
$$\forall k \in [1, 2, \cdots, K]$$

$$(s_{o_{k,i,j}}^t, e_{o_{k,i,j}}^t) \cup (s_{o_{k,i,q}}^t, e_{o_{k,i,q}}^t) = \varnothing \tag{4}$$
$$j \neq q$$

$$(s_{o_{k,i,j}}^t, e_{o_{k,i,j}}^t) \cup (s_{o_{u,v,w}}^t, e_{o_{u,v,w}}^t) = \varnothing$$
$$M_{k,i,j} = M_{u,v,w} \tag{5}$$

$$W_{l,t} < L_{l,t}^W \tag{6}$$

$$s_{o_{k,i,j}}^t > 0 \tag{7}$$

$$s_{o_{k,i,j}}^t = e_{o_{k,i,j-1}}^t + r_{o_{k,i,j-1}}^t \tag{8}$$

上述公式中,式(1)表示目标函数为最小化设备平均负载;式(2)表示目标函数为最小化总加工时间;式(3)表示插单任务需设备数目在可用设备数  $N_{\text{num}}$  内;式(4)表示同一工件在同一时间只能进入一道工序;式(5)表示同一设备在同一时间只能加工一道工序;式(6)表示设备负荷约束,安排在设备上的总负荷不得超过该设备的最大负荷;式(7)定义了工件的首道工序开始时间;式(8)表示本道工序开始时间  $s_{o_{k,i,j}}^t$  在前道工序完成时间  $e_{o_{k,i,j-1}}^t$  后,需考虑设备间运输时间  $r_{o_{k,i,j-1}}^t$ .

2 双系统强化学习方法

为了实现柔性作业车间的动态调度,解决调度过程中的多阶段方案生成优化,本文提出了知识图谱驱动的车间动态调度过程中感知-认知双系统框

架。如图 1 所示,系统 I 的感知是将各类型数据进行表示与集成,将调度信息转变成多维矩阵的形式分别表示资源配置和工序排序,输入到系统 II。在紧急任务的触发下,系统 II 的认知利用 DL-DQN 强化学习方法进行车间排产优化——资源配置智能体以

最小化设备平均负载为目标,首先进行该工件所需人员、设备、物料等资源匹配形成资源社群;工序排序智能体以最小化完工时间为目标,将资源社群以工序为单位实现工件顺序排产计划,分步进行调度策略生成,提高加工流的生产连续性和效率。

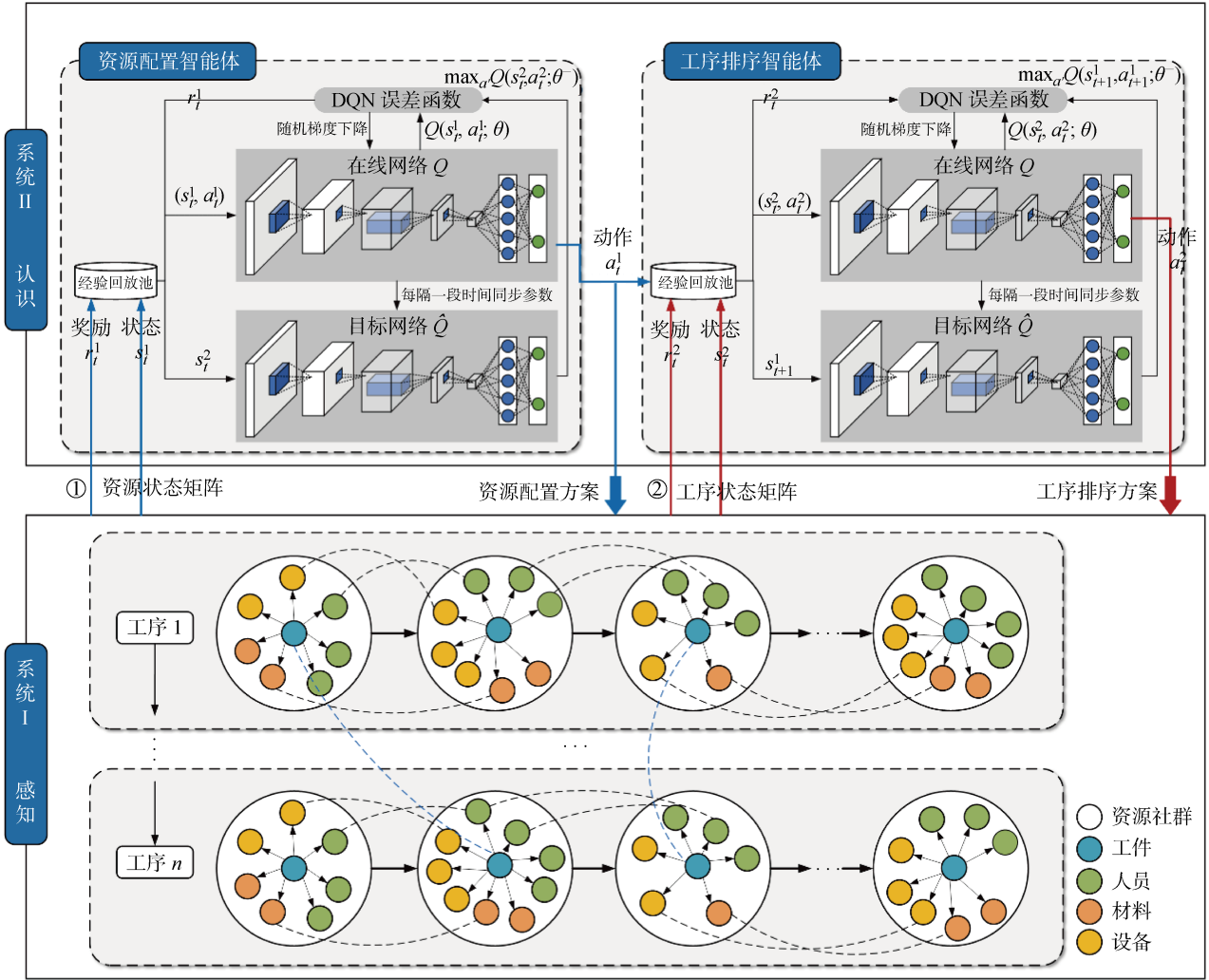


图 1 双系统强化学习框架

Fig. 1 Dual-system reinforcement learning framework

本文将 workflow 调度过程看作是一个马尔科夫决策过程,调度目标为最小化设备总负载和最小化总完工时间,分别将优化目标抽象成两个智能体,智能体之间形成继承关系可以观察到对方的动作和奖励。在考虑双环强化过程的前提下,外环强化过程中资源配置状态处于一个状态的长期最优价值为  $Q(s_t^1, s_t^1; \theta)$ , 执行某个动作之后带来长期最优价值为  $\max_a Q(s_t^2, s_t^2; \theta^-)$ ; 内环强化过程中工序排序状态处于一个状态的长期最优价值为  $Q(s_t^2, s_t^2; \theta)$ , 执行某个动作之后带来长期最优价值为  $\max_a Q(s_{t+1}^1, s_{t+1}^1; \theta^-)$ , 在强化学习算法中,参数  $\theta$  通过迭代最小化损失函数来更新。

2.1 感知-调度参数表示与多维信息矩阵集成

利用知识图谱提供的调度参数,根据车间动态调度的两个子问题,分别将所需数据转化成多维信息矩阵输入到系统的强化学习模型中。

资源配置所需的数据为离散型数据,  $\mathbf{R}$  为资源信息矩阵,每种任务需要由多位工人在多台设备上采用不同的资源进行加工,每种资源包含多种类型:

$$\mathbf{R}_{k,i,j} = \{M_{k,i,j}, P_{k,i,j}, S_{k,i,j}\} \quad (9)$$

$(k = 1, 2, \dots, K; i = 1, 2, \dots, n; j = 1, 2, \dots, m; \mathbf{R}_{k,i,j} \in \mathbf{R}; M_{k,i,j} \in \mathbf{M}; P_{k,i,j} \in \mathbf{P}; S_{k,i,j} \in \mathbf{S})$

工序排序所需的数据由工艺序列和加工时间两组数据组成,  $\mathbf{Q}$  为工序排序信息矩阵,各类型数据构

建出排序涉及相关调度参数,以某时间段工位数据建模表示如下:

$$Q = \{O, T\}$$

(10)

$O$  为工艺参数矩阵,即任务加工工艺方法. 加工车间每种设备组只能对应加工某一类型工艺:

$$O = \{s_{i,j} \mid s_{i,j} = G_1, G_2, \dots, G_l\}$$

(11)

$$(i = 1, 2, \dots, n; j = 1, 2, \dots, m; l = 1, 2, \dots, L)$$

$T$  为时间参数矩阵,包括当前任务在某工序加工的开始时间  $s_{o_k,i,j}^t$ 、结束时间  $e_{o_k,i,j}^t$  和不同工件间的转移时间  $r_{o_k,i,j}^t$ .

$$T = \{t_{i,j} \mid t_{i,j} > 0\} = \{(s_{o_k,i,j}^t, e_{o_k,i,j}^t, r_{o_k,i,j}^t) \mid$$

$$(s_{o_k,i,j}^t > 0, e_{o_k,i,j}^t > 0, r_{o_k,i,j}^t > 0)\}$$

(12)

$$(k = 1, 2, \dots, K; i = 1, 2, \dots, n; j = 1, 2, \dots, m)$$

2.2 认知-DL-DQN 调度算法

对感知数据利用 DL-DQN 算法实现认知应用. 图 2 为本文提出的柔性作业车间动态调度流程. 依据批产任务生产方案流程,对插单任务进行紧急程度的判别后分别应用不同的调度原则,当插单任务进入认知调度实现策略优化时,资源配置智能体形成以工件为核心的资源社群配置,工序排序智能体实现单个工序上的工件排序优化.

本文提出的方法不同于传统调度优化方法,其

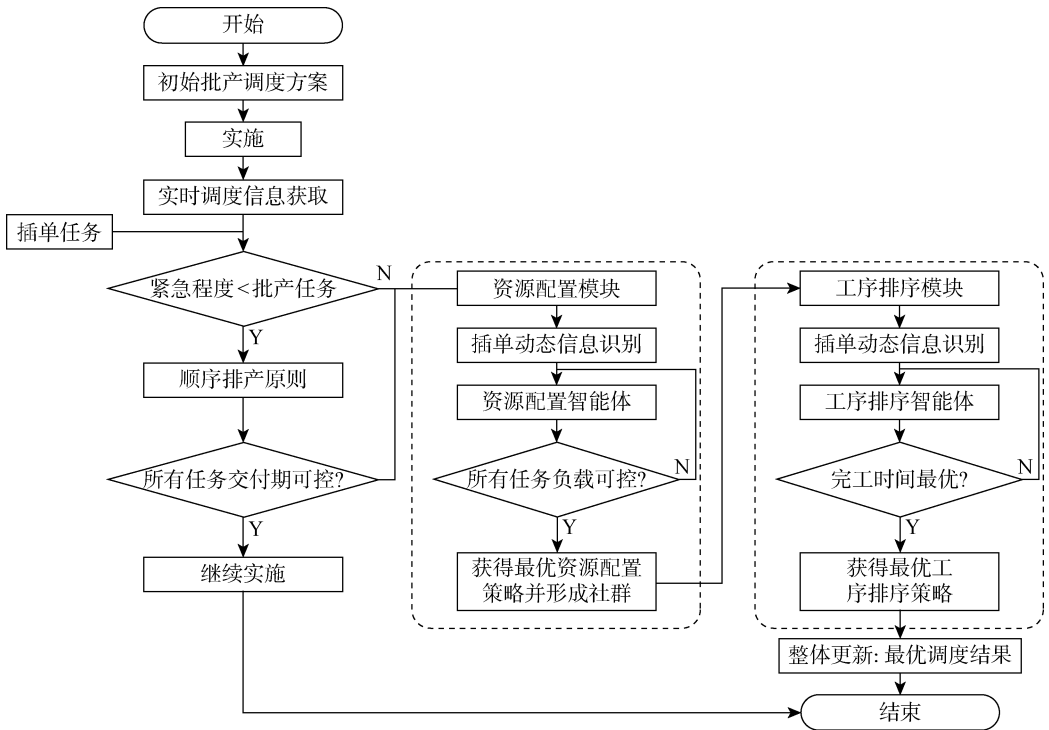


图 2 车间动态调度流程

Fig. 2 Workshop dynamic scheduling process

基于优化目标赋予每个智能体不同的功能,然后与多个子环境交互,根据知识图谱提供的各类型信息和应用场景的协同反馈逐步调整策略实现优化. 图 3 为 DL-DQN 强化学习方法框架, A1 是资源配置智能体, A2 是工序排序智能体. 外环强化由 A1 根据当前加工资源的参数状态  $s_i^1$  对工件的人员、物料和设备等资源进行配置,生成策略  $a_i^1$  到调度场景,对资源配置策略进行以设备平均负载为目标的评估然后将奖励  $r_i^1$  返回到 A1. 资源配置策略状态与当前加工场景的工序排序状态共同作为状态  $s_i^2$  激励 A2 进行内环强化,内环强化的 A2 生成工序排序策略  $a_i^2$  到调度场景,对工序排序策略进行以最小化完

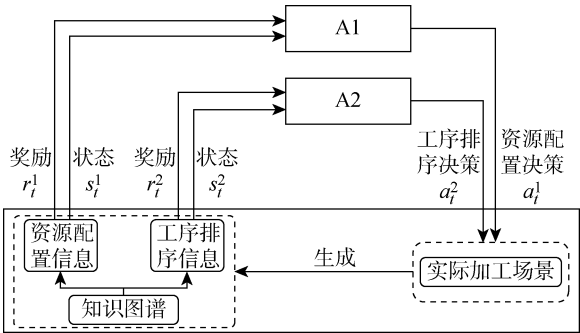


图 3 DL-DQN 强化学习方法框架

Fig. 3 Framework of DL-DQN reinforcement learning algorithm

工时间为目标的评估然后将奖励  $r_t^2$  反馈到 A2,至此完成一个 DL-DQN 强化学习循环.

(1) 调度状态空间.

多智能体在进行车间调度时依赖车间的状态感知信息,分别设计不同的状态空间矩阵.结合车间调度约束和优化目标设计状态矩阵  $F$ ,车间调度任务主要包括资源配置和工序排序两个阶段,因此用  $F=[F^1 \ F^2]$  对车间状态进行描述,  $F^1=[f_1^1 \ f_2^1]$  为车间资源状态,  $F^2=[f_1^2 \ f_2^2]$  为工件加工工艺和

加工时间状态,  $f_1^1=[f_{1,1}^1 \ f_{1,2}^1 \ f_{1,3}^1]$ ,  $f_2^1=[f_{1,1}^2 \ f_{1,2}^2]$  分别表示批产任务的两阶段状态,  $f_2^1, f_2^2$  分别表示插单任务的两阶段状态,  $f_{1,1}^1=[f_{1,1,1}^1]$  表示人员状态的特征向量,  $f_{1,2}^1=[f_{1,2,1}^1]$  表示物料状态的特征向量,  $f_{1,3}^1=[f_{1,3,1}^1 \ f_{1,3,2}^1]$  表示设备状态的特征向量,  $f_{1,1}^2=[f_{1,1,1}^2 \ f_{1,1,2}^2]$  表示工艺状态的特征向量,  $f_{1,2}^2=[f_{1,2,1}^2 \ f_{1,2,2}^2 \ f_{1,2,3}^2 \ f_{1,2,4}^2]$  表示时间状态的特征向量.批产任务加工状态的参数含义及表达式如表 2 所示,插单任务所需

表 2 调度状态参数  
Tab. 2 Parameters of scheduling status

调度状态	参数类型	表达式	含义
资源配置状态	人员状态 $f_{1,1}^1$	$f_{1,1,1}^1=P_{k,i,j}$	人员类型
	物料状态 $f_{1,2}^1$	$f_{1,2,1}^1=S_{k,i,j}$	物料类型
	设备状态 $f_{1,3}^1$	$f_{1,3,1}^1=M_{k,i,j}$	设备类型
		$f_{1,3,2}^1=B_{M_{k,i,j}}^{I_{k,i}}=\begin{cases} 0, & \text{批产任务 } k \text{ 中的工件 } i \text{ 不在设备 } M_{k,i,j} \text{ 上加工} \\ 1, & \text{批产任务 } k \text{ 中的工件 } i \text{ 在设备 } M_{k,i,j} \text{ 上加工} \end{cases} \quad (M_{k,i,j} \in M)$	设备状态
工序排序状态	工艺状态 $f_{1,1}^2$	$f_{1,1,1}^2=o_{k,i,j}$	工艺状态
		$f_{1,1,2}^2=G_{k,i,j}$	设备组状态
	时间状态 $f_{1,2}^2$	$f_{1,2,1}^2=s_{o_{k,i,j}}^t$	开始时间
		$f_{1,2,2}^2=e_{o_{k,i,j}}^t$	结束时间
		$f_{1,2,3}^2=r_{o_{k,i,j}}^t$	运输时间
		$f_{1,2,4}^2=D_k^p$	交付期

参数与批产任务相同.

(2) 调度动作空间.

资源配置和工序排序智能体根据机器的加工状况以及工件的生产属性,分别为每一个工件设置调度优先值,本文分别针对两个智能体设置不同的决策动作空间.

资源配置决策空间:资源配置任务是对插单任务进行人员、设备、物料等资源的配置,合理选择能够最小化设备平均负载的最优资源,决策空间定义如表 3 所示.

表 3 资源配置决策动作  
Tab. 3 Decision-making action of resource allocation

决策方法	描述
决策 1	若人员、设备当前工作环境中空闲且资源配料充足,选择某工件所需人员、设备、物料等资源,然后将其组合为资源配置方案.
决策 2	若人员、设备当前工作环境中忙碌且资源配料不充足,分别判断人员已工作工时与技能参数、设备的负载量和资源的存储量,在满足设备负载量最优的情况下将其组合为资源配置方案.

工序排序决策空间:工序排序是将配置好的资源社群,以最小化完工时间为目标分别在每个工位上进行排序,决策空间定义如表 4 所示.

(3) 调度奖励函数.

通过设置合理的奖励机制,保证算法在多智能体设置下的收敛性,函数设计如下.

资源配置智能体奖励函数  $R_1$ .

首先定义设备繁忙、空闲的示性函数  $\eta_{M_{l,t}}(t_a)$ ,表示为

$$\eta_{M_{l,t}}(t_a)=\begin{cases} 0, & \text{在 } t_a \text{ 时刻机器 } M_{l,t} \text{ 处于空闲状态} \\ -1, & \text{在 } t_a \text{ 时刻机器 } M_{l,t} \text{ 处于繁忙状态} \end{cases} \quad (13)$$

$r_u^1$  表示第  $u$  步决策时刻的立即奖励,公式如下:

$$r_u^1=\frac{1}{N_M}\sum_{l=1}^L\sum_{t=1}^T\int_{t_{u-1}}^{t_u}\eta_{M_{l,t}}(t_a)dt_a \quad (14)$$

式中:  $N_M$  为设备总数;  $t_{u-1}$  和  $t_u$  为第  $u-1$  和第  $u$  个决策时刻;  $r_u^1$  为系统在  $t_u$  时刻获得的关于设备平均负载的即时奖励.该奖励的意义在于:最小化设备平均负载等价于最大化运行一次试验获得的总奖励  $R_1$ .

表 4 工序排序决策动作

Tab. 4 Decision-making action of process sequencing

符号	描述	量化方式
FIFO	先到先加工优先规则	$a_i^2 = \min r_{k,i}(r_{k,i} \text{ 为释放时间})$
SPT	工序加工时间最短优先规则	$a_i^2 = \min \sum_{k=1}^K \sum_{j=1}^n (e_{o_{k,i,j}}^t - s_{o_{k,i,j}}^t + r_{o_{k,i,j}}^t)$
EDD	交货期最早加工优先规则	$a_i^2 = \min D_k^p$
SL	松弛时间最短优先规则	$a_i^2 = \min (D_k^p - x - \sum_{j=1}^m B_{k,i}^T)(x \text{ 为当前时间})$
SRPT	剩余加工时间最长优先规则	$a_i^2 = \max \sum_{j=j'}^m B_{k,i}^T(j' \text{ 为当前工序})$

$$R_1 = \sum_{u=1}^U r_u^1 = \frac{1}{N_M} \sum_{u=1}^U \sum_{l=1}^L \sum_{t=1}^T \int_{t_{u-1}}^{t_u} \eta_{M_{l,t}}(t_a) dt_a =$$
$$\frac{1}{N_M} \sum_{l=1}^L \sum_{t=1}^T \int_0^{T_{l,t}} \eta_{M_{l,t}}(t_a) dt_a =$$
$$- \frac{1}{N_M} \sum_{l=1}^L \sum_{t=1}^T (e_{l,t}^t - s_{l,t}^t) =$$
$$- \frac{1}{N_M} \sum_{l=1}^L \sum_{t=1}^T W_{l,t} \tag{15}$$

$$\max(\sum_{u=1}^U r_u^1) \Leftrightarrow \max - \frac{1}{N_M} \sum_{l=1}^L \sum_{t=1}^T W_{l,t} \Leftrightarrow$$
$$\min\left(\frac{1}{N_M} \sum_{l=1}^L \sum_{t=1}^T W_{l,t}\right) \tag{16}$$

由式(15)和(16)可知,资源配置阶段的最大化奖励  $R_1$  等价于最小化设备平均负载,证明定义的奖励函数有效。

工序排序智能体奖励函数  $R_2$ 。

首先定义  $t_a$  时刻第  $k$  个任务的第  $i$  个工件中第  $j$  道工序状态的示性函数  $\mu_{o_{k,i,j}}(t_a)$ , 即:

$$\mu_{o_{k,i,j}}(t_a) =$$
$$\begin{cases} 0, & \text{在 } t_a \text{ 时刻第 } k \text{ 个任务的第 } i \text{ 个} \\ & \text{工件中第 } j \text{ 道工序没有开始加工} \\ -1, & \text{在 } t_a \text{ 时刻第 } k \text{ 个任务的第 } i \text{ 个} \\ & \text{工件中第 } j \text{ 道工序已经开始加工} \end{cases} \tag{17}$$

$r_u^2$  表示第  $u$  步决策时刻的立即奖励, 公式如下:

$$r_u^2 = \sum_{k=1}^K \sum_{i=1}^n \int_{t_{u-1}}^{t_u} \mu_{o_{k,i,j}}(t_a) dt_a \tag{18}$$

式中:  $t_{u-1}$  和  $t_u$  为第  $u-1$  和第  $u$  个决策时刻;  $r_u^2$  为系统在  $t_u$  时刻获得的关于总完工时间的即时奖励。该奖励的意义在于最小化总完工时间等价于最大化运行一次试验获得的总报酬  $R_2$ 。假设  $D_{k,i,j}^p$  为第  $k$  个任务的第  $i$  个工件中第  $j$  道工序的交付时间。

$$R_2 = \sum_{u=1}^U r_u^2 = \sum_{u=1}^U \sum_{k=1}^K \sum_{i=1}^n \sum_{j=1}^m \int_{t_{u-1}}^{t_u} \mu_{o_{k,i,j}}(t_a) dt_a =$$

$$\sum_{k=1}^K \sum_{i=1}^n \sum_{j=1}^m \int_0^{D_{k,i,j}^p} \mu_{o_{k,i,j}}(t_a) dt_a =$$
$$- \sum_{k=1}^K \sum_{i=1}^n \sum_{j=1}^m (S_{k,i}^T +$$
$$B_{M_{k,i,j}}^{J_{k,i}} (e_{o_{k,i,j}}^t - s_{o_{k,i,j}}^t + r_{o_{k,i,j}}^t)) =$$
$$- \sum_{k=1}^K \sum_{i=1}^n (S_{k,i}^T +$$
$$\sum_{j=1}^m B_{M_{k,i,j}}^{J_{k,i}} (e_{o_{k,i,j}}^t - s_{o_{k,i,j}}^t + r_{o_{k,i,j}}^t)) =$$
$$- \sum_{k=1}^K \sum_{i=1}^n B_{k,i}^T \tag{19}$$

$$\max(\sum_{u=1}^U r_u^2) \Leftrightarrow \max - \sum_{k=1}^K \sum_{i=1}^n B_{k,i}^T \Leftrightarrow$$
$$\min(\sum_{k=1}^K \sum_{i=1}^n B_{k,i}^T) \tag{20}$$

由式(19)和(20)可知,工序排序阶段的最大化奖励  $R_2$  等价于最小化总完工时间,证明定义的奖励函数有效。

为双目标奖励函数赋予权值转变为总奖励函数  $R$ , 如下式:

$$R = \mu_1 R_1 + \mu_2 R_2 \tag{21}$$

式中:  $\mu_1$ 、 $\mu_2$  分别为两个目标的权值,  $\mu_1 + \mu_2 = 1$ 。将双目标优化转换为单目标优化的奖励函数, 通过控制权值的大小, 决策者可依据目标的侧重情况人为设定。例如, 当调度环境要求考虑尽快完工, 可将工序排序智能体奖励函数  $R_2$  的权值设置为 0.7, 资源配置智能体奖励函数  $R_1$  的权值设置为 0.3。

(4) 模型更新。

本文模型在 DQN 基础上改进, 两个智能体在训练时, 在单位时间内, 资源配置智能体根据知识图谱提供的资源相关离散型知识生成资源配置决策动作, 随着状态变化以社群的形式传递给工序排序智能体, 下一个排序动作以每一设备组为单位, 对资源社群进行任务加工序列的安排。调度的两阶段过程

不断更新迭代,直至任务完成,采用梯度下降法更新参数,实现策略函数的优化,算法的伪代码如算法 1 所示.

算法 1:DL-DQN 多智能体强化学习算法

1. 初始化记忆池  $D^1$ 、 $D^2$ ;容量  $N$

2. 用随机权重参数  $\theta^1$ 、 $\theta^2$  初始化函数  $Q^1$ 、 $Q^2$

3. 用随机权重参数  $\theta^{1-}$ 、 $\theta^{2-}$  初始化函数  $\hat{Q}^1$ 、 $\hat{Q}^2$

4. For 迭代次数  $N = 1, M$  do

5.   重置资源配置计划并清除结果,得到初始状态  $s_t^1$

6.   For  $t=0, T$  do

7.     选择资源配置策略  $a_t^1$  并在调度环境中执行操作  $a_t^1$

8.     得到立即奖励  $r_t^1$  和下一个状态  $s_{t+1}^2$

9.     将经历  $(s_t^1, a_t^1, r_t^1, s_{t+1}^2)$  存放到记忆池  $D^1$

10.   For  $N=1, M$  do

11.     重置工序排序计划,继承上一个状态  $s_t^2$

12.     For  $t=0, T$  do

13.       选择工序排序策略  $a_t^2$  并从规则库  $Z = \{\text{FIFO}, \text{SPT}, \text{EDD}, \text{SL}, \text{SRPT}\}$  中选择一个规则  $a_t^2$  进行工序排序

14.       得到立即奖励  $r_t^2$  和下一个状态  $s_{t+1}^1$

15.       将经历  $(s_t^2, a_t^2, r_t^2, s_{t+1}^1)$  存放到记忆池  $D^2$

16.       从记忆池  $D^2$  中随机选择一个转换样本数据,用  $(s_j^2, a_j^2, r_j^2, s_{j+1}^1)$  表示

17.        $y_i^2 = \begin{cases} r_j^2, & N \text{ 在 } j+1 \text{ 步是结束状态} \\ r_j^2 + \gamma^2 \max_{a^2} \hat{Q}^2(s_{j+1}^2, a^2; \theta^{2-}), & \text{其他} \end{cases}$

18.       采用梯度下降法  $(y_i^2 - Q^2(s_j^2, a_j^2; \theta^2))$  更新参数  $\theta^2$

19.       每过  $C$  步迭代更新目标网络  $\hat{Q}^2 = Q^2$

20.     End for

21.   End for

22.   从记忆池  $D^1$  中随机选择一个转换样本数据,用  $(s_j^1, a_j^1, r_j^1, s_{j+1}^2)$  表示

23.    $y_i^1 = \begin{cases} r_j^1, & N \text{ 在 } j+1 \text{ 步是结束状态} \\ r_j^1 + \gamma^1 \max_{a^1} \hat{Q}^1(s_{j+1}^1, a^1; \theta^{1-}), & \text{其他} \end{cases}$

24.   采用梯度下降法  $(y_i^1 - Q^1(s_j^1, a_j^1; \theta^1))$  更新参数  $\theta^1$

25.   每过  $C$  步迭代更新目标网络  $\hat{Q}^1 = Q^1$

26.   End for
27. End for
28. 输出:  $Q^1$ 、 $Q^2$

3 案例与应用

对提出的双系统强化学习方法进行模型训练和实验.首先,对仿真参数进行说明并进行方法对比;然后利用某航天所薄壁壳体加工过程中紧急插单问题进行实例测试,并进行分析.

3.1 模型训练与对比

3.1.1 测试数据 本研究根据 Brandimarte<sup>[28]</sup>提出的方法生成一系列柔性作业车间动态调度问题标准算例,将本文问题进行简化,每种加工任务的工序数相同,根据不同的工序随机选择加工设备,然后在工序加工时间区间内随机生成对应时间(见表 5).

表 5 测试算例参数

Tab. 5 Parameters of test cases

算例	工件数目	工序数	设备数	工序加工时间/h
MK01	6	6	4	[1, 10]
MK02	6	6	6	[3, 8]
MK03	6	8	6	[2, 6]
MK04	8	6	4	[1, 10]
MK05	8	8	6	[2, 9]
MK06	10	6	4	[1, 10]
MK07	10	6	6	[3, 8]
MK08	10	8	6	[2, 6]
MK09	12	6	4	[1, 10]
MK10	12	8	6	[2, 9]

为验证本文提出的 DL-DQN 算法效果,将随机算法(Random Algorithm, RA)、GA、DQN 作为对比算法进行分析比较.

3.1.2 时间复杂度分析 首先,将本文方法和对比算法的时间复杂度分析如下.

(1) RA<sup>[29]</sup>.以先到先加工原则为约束,随机选择设备需加工工件并进行资源分配,设找到可行解需要的迭代次数为  $T^{\text{RA}}$ ,因此 RA 的时间复杂度可近似描述为  $O(T^{\text{RA}}K)$ .

(2) GA<sup>[30]</sup>.遗传算法一般很难收敛到全局最优解,因此算法的停止根据设置的迭代次数,设 GA 的最大迭代次数为  $T^{\text{GA}}$ ,种群数量为  $\lambda$ ,则 GA 的时间复杂度可近似描述为  $O(T^{\text{GA}}\lambda K)$ .

(3) DQN.在 DQN 网络的运行阶段,运行回合



数为  $T^{\text{DQN}}$ , 每回合中的步长为  $n_t$ , 因此在运行阶段的时间复杂度为  $O(T^{\text{DQN}} n_t K)$ . 本文提出的 DL-DQN 算法的时间复杂度与 DQN 相同. 表 6 为时间复杂度对比.

表 6 时间复杂度对比

Tab. 6 Comparison of time complexity

算法名称	时间复杂度
RA	$O(T^{\text{RA}} K)$
GA	$O(T^{\text{GA}} \lambda K)$
DQN	$O(T^{\text{DQN}} n_t K)$
DL-DQN	$O(T^{\text{DL-DQN}} n_t K)$

由表可见, 本文算法和所提算法的时间复杂度和任务数成线性关系, 不同的算法涉及参数不同会影响时间复杂度.

**3.1.3 模拟结果** 本研究以 MK04 算例进行实验评估, 构建作业车间环境, 并使用 DL-DQN 算法进行柔性作业车间调度, 在迭代过程中记录参数变化, 与 DQN 算法进行对比. 训练过程参数变化如图 4 所示.

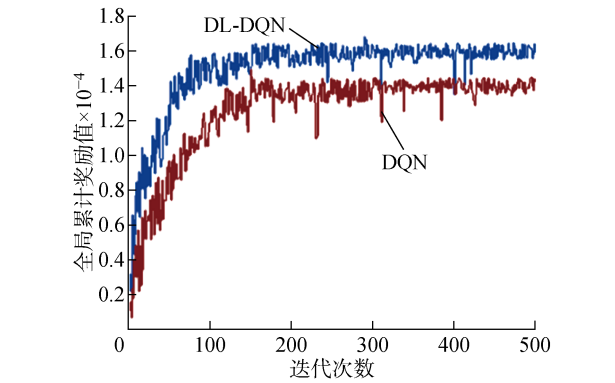


图 4 MK04 算例模拟参数收敛曲线

Fig. 4 Convergence curve of simulation parameters for Case MK04

根据测试算例 MK01~MK10, 分别用本文方法和对比方法求解总优化目标, 结果如图 5 所示.

由图 4 可见 DL-DQN 算法训练过程的全局累积奖励表现出良好的上升趋势, 收敛速度较 DQN 模型快, 优化过程优于 DQN 模型, 且得到的解决方案更好, 更能适应柔性作业车间调度的动态性. 进一步, 通过本文算法和对比算法的求解应用 (见图 5), 所提方法的设备平均负载与总完工时间的加权值均小于对比算法. 整体看来, DL-DQN 方法效果较优, 验证了本文所提方法的有效性.

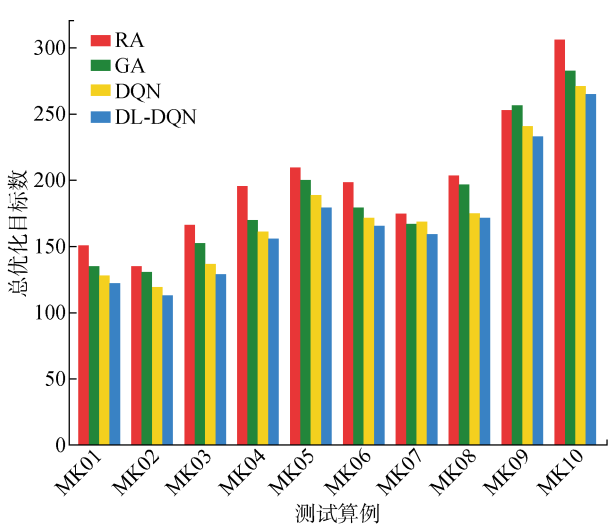


图 5 测试算例不同算法的总优化目标

Fig. 5 Total optimization goals of different algorithms for test cases

3.2 实例测试

**3.2.1 加工车间任务信息表** 利用某动力所固体火箭发动机薄壁壳体生产模拟数据进行实例测试以验证所提方法的有效性. 薄壁壳体的加工具有产研并存的特性, 整理航天所生产原始数据后得到设备类型与相关信息如表 7 所示, 不同的设备组对应不同的工序.

表 7 车间工序对应多功能设备组

Tab. 7 Optional multifunctional equipment group for workshop

工序	设备组	设备组序列
固溶	固溶炉组	G <sub>1</sub>
旋压	旋压机组	G <sub>2</sub>
退火	退火炉组	G <sub>3</sub>
时效	时效炉组	G <sub>4</sub>
粗加工/精加工	数控机床组/加工中心组	G <sub>5</sub>
电子束组焊/激光焊	焊机组	G <sub>6</sub>
油淬	油淬炉组	G <sub>7</sub>
氮气淬	氮气淬炉组	G <sub>8</sub>
回火	回火炉组	G <sub>9</sub>

案例以薄壁壳体批产任务加工过程中紧急订单触发任务为分析对象, 在设备信息基础上, 设置 6 个任务的工艺约束、加工时间、加工数量和交货期, 其中包含 4 个批产任务 ( $J_1 \sim J_4$ ) 与 2 个插单任务 ( $J_5 \sim J_6$ ). 生产任务工序信息表和时间表如表 8 和表 9 所示, 其中“空白”表示无此项.

表 8 生产任务工序信息表

Tab. 8 Process information of production task

任务	工序 1	工序 2	工序 3	工序 4	工序 5	工序 6	工序 7	工序 8
J <sub>1</sub>	固溶	旋压	时效	粗加工	焊接	氮气淬	回火	精加工
J <sub>2</sub>	旋压	退火	粗加工	焊接	氮气淬	回火	精加工	
J <sub>3</sub>	固溶	旋压	粗加工	焊接	油淬	回火	精加工	
J <sub>4</sub>	固溶	旋压	退火	粗加工	焊接	油淬	回火	精加工
J <sub>5</sub>	固溶	旋压	退火	粗加工	焊接	精加工		
J <sub>6</sub>	固溶	旋压	时效	粗加工	焊接	氮气淬	回火	精加工

表 9 生产任务工序时间表

Tab. 9 Process time of production task

任务	工时/h								加工数量	交货时间/h
	工序 1	工序 2	工序 3	工序 4	工序 5	工序 6	工序 7	工序 8		
J <sub>1</sub>	40	15	53	25	19	73	20	17	20	300
J <sub>2</sub>	10	43	24	11	38	13	12		10	200
J <sub>3</sub>	33	33	26	17	103	22	20		18	280
J <sub>4</sub>	13	5	4	5	5	17	4	5	3	300
J <sub>5</sub>	16	8	36	19	9	14			8	120
J <sub>6</sub>	9	3	23	8	6	28	11	10	5	120

3.2.2 感知系统知识表示与集成 根据本文提出的 DL-DQN 算法,分别为资源配置智能体和工序排序智能体提供多维信息矩阵,两阶段调度过程中所需的参数矩阵如图 6 所示。

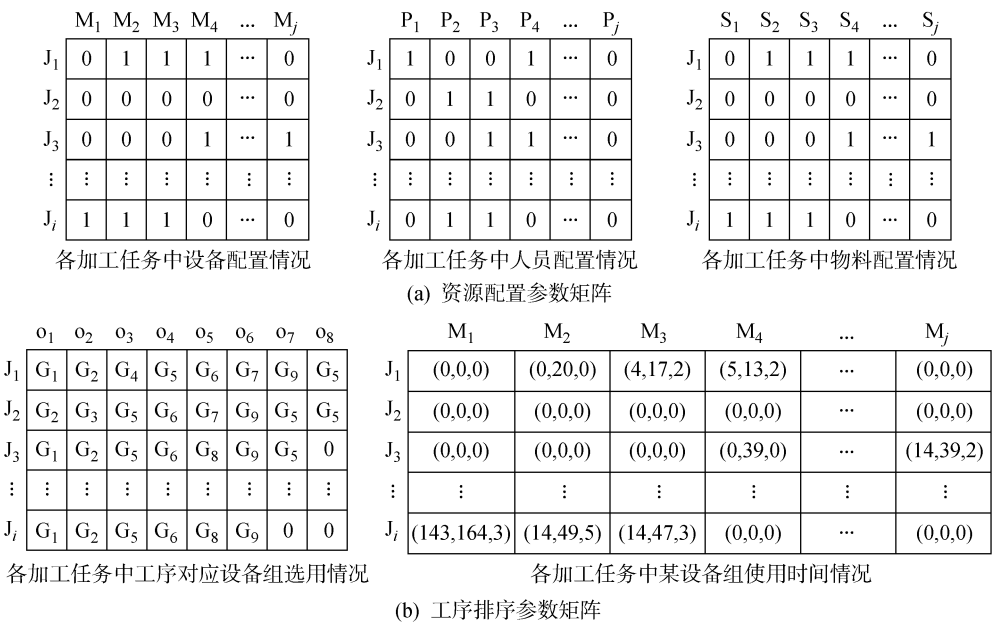


图 6 多维信息矩阵实例化表示

Fig. 6 Instantiated representation of multi-dimensional information matrix

3.2.3 认知系统车间调度策略生成 假设插单任务的触发时间节点位于批产任务初始化后的 20 h 和 80 h 处,针对插单任务导致的柔性作业车间动态调度事件,首先对其进行资源配置,然后以资源社群为单位进行工序排序,将插单任务排产到合适的工序位置。

在资源配置过程中,针对机加工车间的车床工作状态进行全方位的任务配置. 以该车间某一时段的车床加工任务作为分析基础,将车床按照类型划分,每种型号的车床设备各 1 台,其加工功能不同,针对插单过程中导致的设备平均负载变化,车间车床设备相关信息如表 10 所示.

为表明资源配置智能体生成配置策略的合理性,以设备利用率为参考得出调度前后车床设备利用率曲线图(见图 7),可以直观地发现,车间车床利用率呈现整体上升趋势且变化平缓,较少出现某一设备负载突变的情况.

表 10 车间车床相关信息表

Tab. 10 Related information of workshop lathe

设备类别	设备型号
普通卧式车床	CDZ6140-1, CDZ6140-2
	CD6140B-1
数控卧式车床	CK61200W-1
	CK6146A-1
	CK64160-1, CK64160-2, CK64160-3
	CK64250-1
CKD 系列数控车床	CKD6163-1
	CKD6163K-1, CKD6163K-2
	CKD6140S-1
	CKD6180D-1, CKD6180D-2
管螺纹车床	QK1319A-1, QK1319A-2

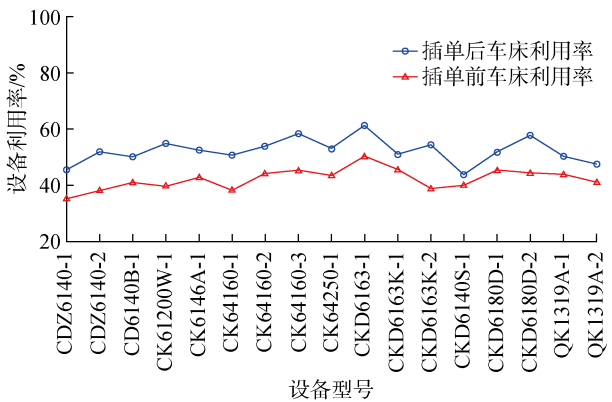


图 7 插单前后车床利用率信息

Fig. 7 Lathe utilization information before and after order insertion

将插单任务触发前后的车间调度模式以甘特图的形式表示出来,图 8 为插单前任务调度方案结果,图 9 为插单后任务调度方案结果. 不同的颜色代表不同任务,同一任务按照时序关系对工序进行区分.

根据插单前后的甘特图得出不同任务交付时间与加工时长的对比. 如表 11 所示(“空白”表示无此项),运用本文方法能够在满足插单任务交付期的前提下,在一定程度上缩短批产任务的总加工时长,特别是调度策略前后  $J_4$  的总加工时间减少 35.08%,交付时间减少 13.67%. 薄壁壳体制造过程中不同的加工工艺所需时间差异大,因此车间中工序排序引起的时间微小变化会更有利于调度. 从设备利用率和加工时间变化来看,本文提出的双系统深度强化学习方法具有一定的可行性.

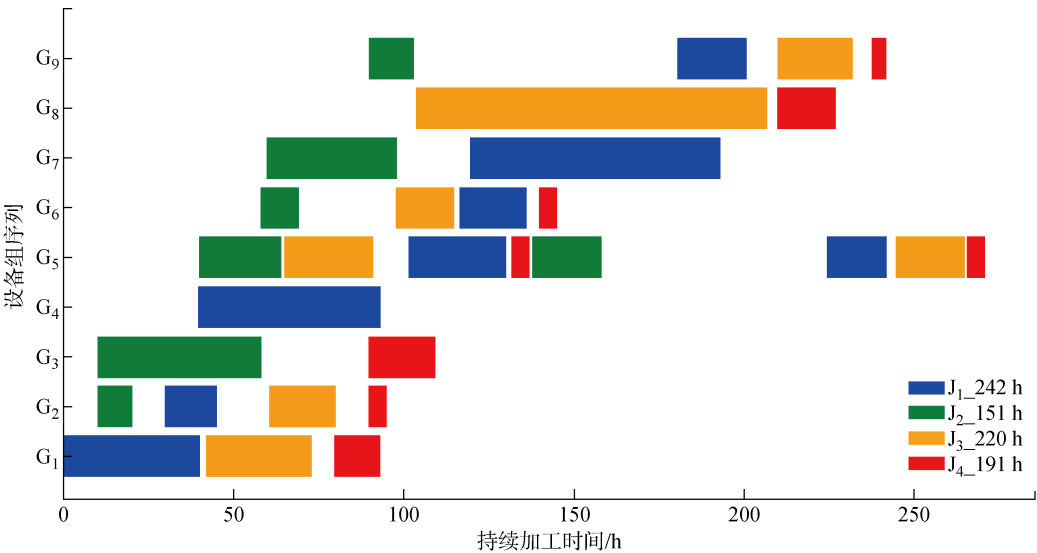


图 8 插单前调度方案

Fig. 8 Scheduling plan before order insertion

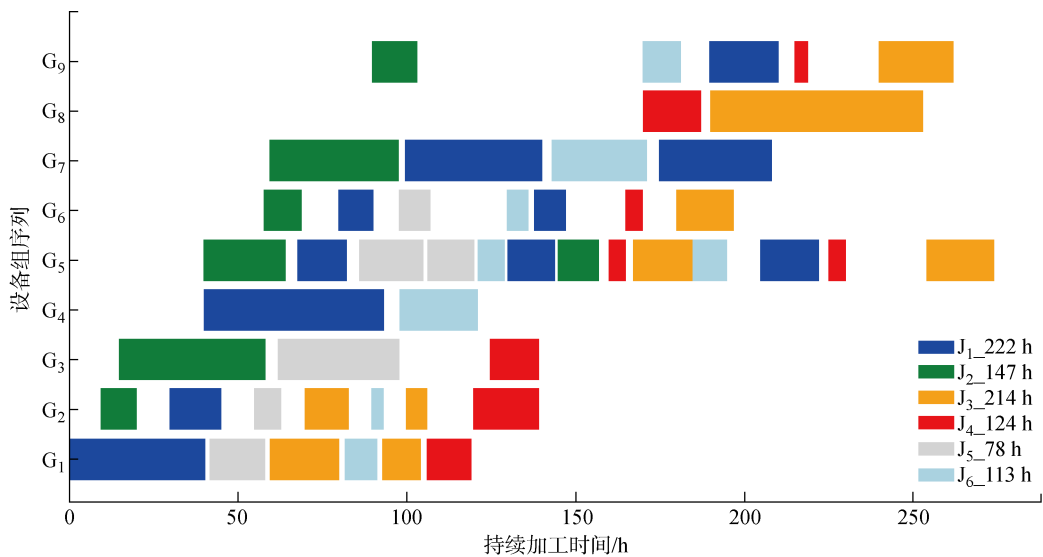


图 9 插单后调度方案

Fig. 9 Scheduling plan after order insertion

表 11 插单前后任务交付时间与加工时长对比表

Tab. 11 Comparison of task delivery time and processing time before and after order insertion

任务	插单前任务 交付期	插单后任务 交付期	插单前任务 加工时长/h	插单后任务 加工时长/h	交付期 变化率/%	加工时长 变化率/%
J <sub>1</sub>	242	222	242	222	6.67	8.26
J <sub>2</sub>	161	157	151	147	2	2.65
J <sub>3</sub>	265	274	220	214	−3.21	2.73
J <sub>4</sub>	271	230	191	124	13.67	35.08
J <sub>5</sub>		100		78		
J <sub>6</sub>		115		113		

4 结论

本文对产研并存的柔性作业车间动态调度问题进行了研究,以最小化设备平均负载和最小化总完工时间为优化目标,提出感知-认知双系统与 DL-DQN 算法融合嵌入的方法,得出以下结论:

- (1) 感知系统将知识图谱提供的数据转化成多维信息矩阵输入到认知系统,认知系统将调度过程分成资源配置和工序排序两部分,分别将两个优化目标抽象成具有继承关系的智能体。
- (2) 采用 DL-DQN 算法,外环利用资源配置智能体生成以工件为单位的资源社群,内环利用工序排序智能体以资源社群为单位进行工序级别的工件排序优化。
- (3) 对比不同的算法和进行某航天所的固体火箭发动机薄壁壳体加工车间的实例测试,证实提出的双环强化学习方法能更好地利用调度知识图谱提

供的参数信息,实现柔性作业车间调度的多目标优化。

目前的研究将设备故障、突发撤单等不确定因素忽略,未来将探究多资源约束情况下的柔性作业车间自适应调度问题。

参考文献:

[ 1 ] JI S X, PAN S R, CAMBRIA E, *et al.* A survey on knowledge graphs: Representation, acquisition, and applications[J]. **IEEE Transactions on Neural Networks and Learning Systems**, 2022, 33(2): 494-514.

[ 2 ] ISSA S, ADEKUNLE O, HAMDI F, *et al.* Knowledge graph completeness: A systematic literature review[J]. **IEEE Access**, 2021, 9: 31322-31339.

[ 3 ] CARVALHO A, CHOUCHE A, LIMA T, *et al.* Cognitive manufacturing in industry 4.0 toward cognitive load reduction: A conceptual framework [J]. **Applied System Innovation**, 2020, 3(4): 55.

- [4] LU Y Q, XU X, WANG L H. Smart manufacturing process and system automation: A critical review of the standards and envisioned scenarios[J]. **Journal of Manufacturing Systems**, 2020, 56: 312-325.
- [5] ZHANG J, DING G F, ZOU Y S, *et al.* Review of job shop scheduling research and its new perspectives under Industry 4.0[J]. **Journal of Intelligent Manufacturing**, 2019, 30(4): 1809-1830.
- [6] 巴智勇, 袁逸萍, 戴毅, 等. 考虑机器故障的作业车间调度方案鲁棒调度方法[J]. **计算机集成制造系统**, 2020, 26(12): 3341-3349.
- BA Zhiyong, YUAN Yiping, DAI Yi, *et al.* Robustness measurement approach of job shop scheduling with machine breakdowns[J]. **Computer Integrated Manufacturing Systems**, 2020, 26(12): 3341-3349.
- [7] 方佳, 陆志强. 考虑设备故障的鲁棒调度计划模板的建模优化[J]. **上海交通大学学报**, 2020, 54(12): 1278-1290.
- FANG Jia, LU Zhiqiang. Modeling and optimization of robust scheduling template considering equipment failure[J]. **Journal of Shanghai Jiao Tong University**, 2020, 54(12): 1278-1290.
- [8] GAO K Z, YANG F J, ZHOU M C, *et al.* Flexible job-shop rescheduling for new job insertion by using discrete jaya algorithm[J]. **IEEE Transactions on Cybernetics**, 2019, 49(5): 1944-1955.
- [9] 王鹏飞. 群智能优化算法及在流水车间调度问题中的应用研究[D]. 长春: 吉林大学, 2019.
- WANG Pengfei. Swarm intelligence optimization algorithm and its application in flow shop scheduling problem[D]. Changchun: Jilin University, 2019.
- [10] ZHOU B, BAO J S, LI J, *et al.* A novel knowledge graph-based optimization approach for resource allocation in discrete manufacturing workshops[J]. **Robotics and Computer-Integrated Manufacturing**, 2021, 71: 102160.
- [11] CHAKRABORTTY R K, RAHMAN H F, RYAN M J. Efficient priority rules for project scheduling under dynamic environments: A heuristic approach [J]. **Computers & Industrial Engineering**, 2020, 140: 106287.
- [12] 蒋小康, 张朋, 吕佑龙, 等. 基于混合蚁群算法的半导体生产线炉管区调度方法[J]. **上海交通大学学报**, 2020, 54(8): 792-804.
- JIANG Xiaokang, ZHANG Peng, LYU Youlong, *et al.* Hybrid ant colony algorithm for batch scheduling in semiconductor furnace operation[J]. **Journal of Shanghai Jiao Tong University**, 2020, 54(8): 792-804.
- [13] 王金凤, 陈璐, 杨雯慧. 考虑设备可用性约束的单机调度问题[J]. **上海交通大学学报**, 2021, 55(1): 103-110.
- WANG Jinfeng, CHEN Lu, YANG Wenhui. A single machine scheduling problem considering machine availability constraints[J]. **Journal of Shanghai Jiao Tong University**, 2021, 55(1): 103-110.
- [14] 杜轩, 潘志成. 聚类差分进化算法求解多目标工艺规划与调度集成问题[J]. **计算机集成制造系统**, 2019, 25(7): 1729-1738.
- DU Xuan, PAN Zhicheng. Clustering and differential evolution algorithm for solving multi-objectives IPPS problem[J]. **Computer Integrated Manufacturing Systems**, 2019, 25(7): 1729-1738.
- [15] 李聪波, 沈欢, 李玲玲, 等. 面向能耗的多工艺路线柔性作业车间分批优化调度模型[J]. **机械工程学报**, 2017, 53(5): 12-23.
- LI Congbo, SHEN Huan, LI Lingling, *et al.* A batch splitting flexible job shop scheduling model for energy saving under alternative process plans [J]. **Journal of Mechanical Engineering**, 2017, 53(5): 12-23.
- [16] PENG C, WU G L, LIAO T W, *et al.* Research on multi-agent genetic algorithm based on tabu search for the job shop scheduling problem [J]. **PLoS One**, 2019, 14(9): e0223182.
- [17] KUNDAKCI N, KULAK O. Hybrid genetic algorithms for minimizing makespan in dynamic job shop scheduling problem[J]. **Computers & Industrial Engineering**, 2016, 96: 31-51.
- [18] SHEN X N, YAO X. Mathematical modeling and multi-objective evolutionary algorithms applied to dynamic flexible job shop scheduling problems[J]. **Information Sciences**, 2015, 298: 198-224.
- [19] WANG Z, ZHANG J H, YANG S X. An improved particle swarm optimization algorithm for dynamic job shop scheduling problems with random job arrivals [J]. **Swarm and Evolutionary Computation**, 2019, 51: 100594.
- [20] 张洁, 张朋, 刘国宝. 基于两阶段蚁群算法的带非等效并行机的作业车间调度[J]. **机械工程学报**, 2013, 49(6): 136-144.
- ZHANG Jie, ZHANG Peng, LIU Guobao. Two-stage ant colony algorithm based job shop scheduling with unrelated parallel machines[J]. **Journal of Mechanical Engineering**, 2013, 49(6): 136-144.
- [21] 周亚勤, 杨长祺, 吕佑龙, 等. 双资源约束的航天结构件车间生产调度方法[J]. **机械工程学报**, 2018, 54(9): 55-63.

ZHOU Yaqin, YANG Changqi, LÜ Youlong, *et al.* Scheduling the production of aerospace structural parts with dual resource constraints[J]. **Journal of Mechanical Engineering**, 2018, 54(9): 55-63.

[22] 汪浩祥, 严洪森, 汪峥. 知识化制造环境中基于双层 Q 学习的航空发动机自适应装配调度[J]. **计算机集成制造系统**, 2014, 20(12): 3000-3010.

WANG Haoxiang, YAN Hongsen, WANG Zheng. Adaptive assembly scheduling of aero-engine based on double-layer Q-learning in knowledgeable manufacturing[J]. **Computer Integrated Manufacturing Systems**, 2014, 20(12): 3000-3010.

[23] WEI Y, PAN L, LIU S J, *et al.* DRL-scheduling: An intelligent QoS-aware job scheduling framework for applications in clouds[J]. **IEEE Access**, 2018, 6: 55112-55125.

[24] WANG Y D, LIU H, ZHENG W B, *et al.* Multi-objective workflow scheduling with deep-Q-network-based multi-agent reinforcement learning[J]. **IEEE Access**, 2019, 7: 39974-39982.

[25] LUO S. Dynamic scheduling for flexible job shop with new job insertions by deep reinforcement learning[J]. **Applied Soft Computing**, 2020, 91: 106208.

[26] HE Z L, TRAN K P, THOMASSEY S, *et al.* Multi-objective optimization of the textile manufacturing process using deep-Q-network based multi-agent reinforcement learning[J]. **Journal of Manufacturing Systems**, 2022, 62: 939-949.

[27] 林时敬, 徐安军, 刘成, 等. 基于深度强化学习的炼钢车间天车调度方法[J]. **中国冶金**, 2021, 31(3): 37-43.

LIN Shijing, XU Anjun, LIU Cheng, *et al.* Crane scheduling method in steelmaking workshop based on deep reinforcement learning[J]. **China Metallurgy**, 2021, 31(3): 37-43.

[28] BRANDIMARTE P. Routing and scheduling in a flexible job shop by tabu search[J]. **Annals of Operations Research**, 1993, 41(3): 157-183.

[29] 喻鹏, 张俊也, 李文璟, 等. 移动边缘网络中基于双深度 Q 学习的高能效资源分配方法[J]. **通信学报**, 2020, 41(12): 148-161.

YU Peng, ZHANG Junye, LI Wenjing, *et al.* Energy-efficient resource allocation method in mobile edge network based on double deep Q-learning[J]. **Journal on Communications**, 2020, 41(12): 148-161.

[30] 牟乃夏, 徐玉静, 李洁, 等. 遗传禁忌搜索算法收敛性和时间复杂度分析[J]. **河南理工大学学报(自然科学版)**, 2018, 37(4): 118-122.

MOU Naixia, XU Yujing, LI Jie, *et al.* Analyses of convergence and time complexity of genetic tabu search algorithm[J]. **Journal of Henan Polytechnic University (Natural Science)**, 2018, 37(4): 118-122.

(本文编辑:孙伟)