



中国科技期刊卓越行动计划项目入选期刊

控制与决策

CONTROL AND DECISION



基于深度强化学习的多人协同混流装配线平衡优化研究

张梅, 田镇遇, 朱金辉, 傅艳霞

引用本文:

张梅, 田镇遇, 朱金辉, 傅艳霞. 基于深度强化学习的多人协同混流装配线平衡优化研究[J]. *控制与决策*, 2024, 39(10): 3395–3404.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2023.0820>

您可能感兴趣的其他文章

Articles you may be interested in

考虑批量分割的单元装配系统工作量平衡联合决策模型

Joint decision model of Seru production system balancing considering lot-splitting
控制与决策. 2021, 36(10): 2517–2527 <https://doi.org/10.13195/j.kzyjc.2020.0167>

铁路集装箱中心站资源分配与作业调度联合优化

Integrating optimization of resource allocation and handling scheduling in railway container terminal
控制与决策. 2021, 36(12): 3063–3073 <https://doi.org/10.13195/j.kzyjc.2020.0597>

基于资源特征融合的复杂型面配合决策方法

Decision-making method for complex surface coordination based on resource feature fusion
控制与决策. 2021, 36(6): 1368–1376 <https://doi.org/10.13195/j.kzyjc.2019.1311>

基于复杂昂贵仿真的体系效能多目标优化

Complex and expensive simulation based multi-objective optimization to system-of-system effectiveness
控制与决策. 2021, 36(3): 589–598 <https://doi.org/10.13195/j.kzyjc.2019.0844>

基于知识粒度特征的多目标粗糙集属性约简算法

Multi objective rough set attribute reduction algorithm based on characteristics of knowledge granularity
控制与决策. 2021, 36(1): 196–205 <https://doi.org/10.13195/j.kzyjc.2019.0490>

基于深度强化学习的多人协同混流装配线平衡优化研究

张梅^{1†}, 田镇遇¹, 朱金辉^{2,3}, 傅艳霞¹

(1. 华南理工大学 自动化科学与工程学院, 广州 510641; 2. 华南理工大学 软件学院, 广州 510006;
3. 华南理工大学 大数据与智能机器人教育部重点实验室, 广州 510006)

摘要: 针对大型设备混流装配过程中的多人协同、多工种等特点, 提出基于双深度 Q 网络 (double deep Q network, DDQN) 的多人协同混流装配线平衡优化算法. 首先以工作站和工人数量、工人与工作站间的负载为优化目标, 建立多人协同混流装配线平衡问题的多目标优化数学模型. 其次, 根据装配过程中生产对象的特征设计状态空间, 并根据启发式规则设计动作空间, 结合优化目标设计奖励函数, 从而将数学模型转化为马尔科夫决策模型. 在此基础上, 对传统 DDQN 算法进行改进, 采用自适应探索概率完成动作决策, 并设计基于工人利用率的解码方法. 最后, 通过混流装配线标准测试实例以及多人协同混流装配线测试实例, 将 DDQN 算法与改进离散水波优化算法和模拟退火算法进行对比, 验证算法的寻优精度以及模型的有效性. 同时, 在车身混流装配实际案例中采用 DDQN 算法进行平衡优化, 验证算法的有效性和实用性.

关键词: 混流装配线; 平衡优化; 多人协作; 强化学习; 多目标优化

中图分类号: TP273

文献标志码: A

DOI: 10.13195/j.kzyjc.2023.0820

引用格式: 张梅, 田镇遇, 朱金辉, 等. 基于深度强化学习的多人协同混流装配线平衡优化研究 [J]. 控制与决策, 2024, 39(10): 3395-3404.

Multi-manned collaborative mixed-model assembly line balancing optimization based on deep reinforcement learning

ZHANG Mei^{1†}, TIAN Zhen-yu¹, ZHU Jin-hui^{2,3}, FU Yan-xia¹

(1. School of Automation and Engineering, South China University of Technology, Guangzhou 510641, China; 2. School of Software Engineering, South China University of Technology, Guangzhou 510006, China; 3. Key Laboratory of Big Data and Intelligent Robot of Ministry of Education, South China University of Technology, Guangzhou 510006, China)

Abstract: Considering the characteristics of assembly process such as multiple workers collaborating, the demand for workers with different skills, and mixed-model assembly, this paper proposes a double deep Q network (DDQN) based algorithm to address a multi-manned cooperation mixed-model assembly line balancing problem. Firstly, a mathematical model for the multi-manned cooperation mixed-model assembly line balancing problem is established with the objectives of optimising the number of workstations and workers, the workload between workers and workstations. Secondly, the state space is designed based on the features of production objects. Meanwhile, the action space is designed using heuristic rules. Besides, the reward function is constructed based on the objectives of the model. As a result, the mathematical model is converted into a Markov decision process model. On this basis, an improved DDQN algorithm with an adaptive exploration probability for action decision-making and a decoding method based on worker utilization rate is developed. Finally, the improved DDQN algorithm is compared with the improved discrete water wave optimization algorithm and the simulated annealing algorithm on standard mixed-model assembly line test cases and multi-manned collaborative mixed-model assembly line test cases to verify the accuracy of the algorithm and the effectiveness of the model. The effectiveness and practicality of the algorithm are also verified by applying it to balance optimization in a practical car body mixed-flow assembly process.

Keywords: mixed-model assembly line; balancing optimization; multi-manned cooperation; reinforcement learning; multi-objective optimization

收稿日期: 2023-06-13; 录用日期: 2023-11-03.

基金项目: 国家重点研发计划项目 (2021YFB3202200).

责任编辑: 王凌.

[†]通讯作者. E-mail: zhangmei@scut.edu.cn.

*本文附带电子附录文件, 可登录本刊官网该文“资源附件”区自行下载阅览.

0 引言

随着市场需求的多样化,小批量的柔性混流装配线逐渐成为制造企业的首选方式. 在一些大型设备装配线上,例如汽车、飞机等,由于产品体积较大,装配工序的复杂度较高,需要一定数量的专业技能工人协同装配一道工序,混流装配线上劳动力资源的合理调配显得尤为重要.

目前,针对混流装配线平衡问题 (mixed-model assembly line balancing problem, MALBP)^[1]的求解方法主要可分为精确算法^[2]、启发式算法^[3]和元启发式算法. 由于MALBP问题的NP-难特性,精确方法难以用于求解大规模问题. 启发式算法主要根据问题特点得到一些启发式规则实现求解,此类算法简单,但存在易陷入局部最优解的缺点. 鉴于精确算法和启发式算法的局限性,近年来元启发式算法逐渐成为国内外学者的研究重点,一些学者考虑了实际装配作业过程中存在的因素,如工作时间的随机性^[4]、最小化能量消耗^[5]等,采用元启发式算法进行求解.

尽管元启发式算法解决了一些装配线平衡问题,但许多算法缺乏对历史信息关注和利用,因此在性能上依然有提升空间. 近年来,深度强化学习 (deep reinforcement learning, DRL) 在围棋、机器人等领域的瞩目成果显示了其强大的学习能力和序列决策能力. 因此,一些学者基于深度强化学习方法在车间调度^[6-8]、路径规划^[9-10]等组合优化问题的研究成果,将DRL方法用于求解装配线调度过程中的相关问题. 赵铭慧等^[11]采用深度Q网络算法对双向装配序列规划问题进行求解; Brammer等^[12]基于近端策略优化的深度强化学习方法,利用启发式信息求解产品的投产序列.

本文针对大型产品在混流装配过程中的多人协同情况以及混流产品的差异性对负载均衡性的影响,将装配环节涉及的工序分配、工人指派问题综合考虑为多人协同混流装配线平衡问题 (multi-manned cooperation mixed-model assembly line balancing problem, MCMALBP),通过对MCMALBP的特点进行分析,提出一种基于深度强化学习的平衡优化方法,能够很好地降低混流装配过程中的生产资源成本,提高装配效率,为企业的装配线规划提供指导.

1 问题分析及建模

1.1 问题分析

与一般MALBP相比,本文研究的MCMALBP具有以下特点:

1) 工序有所需工人类型和数量的要求,即某道工

序能够开始装配的前提是有足够数量且具备装配该工序技能的空闲工人;

2) 每个工人有唯一的工人类型,即一个工人只能负责处理对应类型的工序,不考虑全能型工人;

3) 工序的处理时间与产品类型相关,即不同产品的同一道工序可能有不同的处理时间,但工序所需工人类型和数量与产品类型无关;

4) 工作站上所能容纳的工人数量存在上限.

1.2 符号定义

本文建立的MCMALBP数学模型中涉及的相关索引号、常量、决策变量的符号定义如表1所示.

表1 相关符号定义

符号	描述
M	产品类型集合
N	工序集合
K	工作站集合
W	工人集合
m	产品类型索引, $m \in M$
i, j	工序索引, $i, j \in N$
k	工作站索引, $k \in K$
w	工人索引, $w \in W$
B	一个很大的正整数
G_i	工序 i 在综合优先关系图中的紧前工序集合
ct	装配线的生产节拍
ct_k	工作站 k 的生产节拍
k_{\max}	最大索引工作站对应的编号
AO	可选工序集合
CO	已完成工序集合
mw	工作站上所能容纳的最大工人数量
d_m	第 m 种类型产品的数量
t_{mi}	第 m 种类型产品中工序 i 的处理时间, 若产品不包含工序 i , 则 $t_{mi} = 0$
\bar{t}_i	工序 i 在综合优先关系图中的平均处理时间 $\bar{t}_i = \frac{\sum_{m \in M} t_{mi}}{\sum_{m \in M} d_m}, \forall i \in N$
p_i	工序 i 所需的工人数量
e_i	工序 i 所需的工人类型
\hat{t}_i	工序 i 的混流处理时间, $\hat{t}_i = \frac{\sum_{m \in M} d_m t_{mi}}{\ M\ }$
dev_i	工序 i 的负载差异性指数, $dev_i = \frac{\sum_{m \in M} (d_m \cdot t_{mi} - \hat{t}_i \cdot p_i)}{\sum_{m \in M} d_m}$
c_i	工序 i 的完工时间
t_w	工人 w 的总装配时间
wc_w	工人 w 的完工时间
x_{ik}	工序 i 被分配到工作站 k 上时为1, 其他为0
y_{iw}	工序 i 被分配给工人 w 上时为1, 其他为0
z_{wk}	工人 w 被指派到工作站 k 上时为1, 其他为0
o_{ij}	工序 i 和工序 j 被分配给同一个工人, 且工序 i 在工序 j 之前完成装配时为1, 其他为0

1.3 数学模型

本文以工作站数量和工人数量的综合指数 f_1 、工人负荷平滑度 f_2 、工作站间混流产品的负载差异性平滑度 f_3 为优化目标,采用加权法得到 MCMALBP 的综合优化目标函数以及约束条件如下:

$$F = \min(\alpha_1 \cdot g_1 \cdot f_1 + \alpha_2 \cdot g_2 \cdot f_2 + \alpha_3 \cdot g_3 \cdot f_3); \quad (1)$$

$$f_1 = \min(w_1 \cdot \|K\| + w_2 \cdot \|W\|); \quad (2)$$

$$f_2 = \min\left(\sqrt{\frac{\sum_{w \in W} (t_w - ct)^2}{\|w\|}}\right); \quad (3)$$

$$f_3 = \min\left(\sqrt{\frac{\sum_{k \in X} \left(\sum_{i \in N} (x_{ik} \cdot dev_i) - \frac{\sum_{i \in N} dev_i}{\|K\|}\right)^2}{\|K\|}}\right). \quad (4)$$

s.t.

$$\sum_{k \in K} x_{ik} = 1, \forall i \in N; \quad (5)$$

$$c_i \leq c_j - \bar{t}_j + B \cdot \left(\sum_{k \in X} (k \cdot x_{jk}) - \sum_{k \in K} (k \cdot x_{ik})\right), \quad \forall j \in N, i \in G_j; \quad (6)$$

$$c_i \leq c_j - \bar{t}_j + B \cdot (1 - o_{ij}), \forall i, j \in N, i \neq j; \quad (7)$$

$$\bar{t}_i \leq c_i \leq ct, \forall i \in N; \quad (8)$$

$$\sum_{i \in N} (\bar{t}_i \cdot y_{iw}) \leq t_w \leq wc_w, \forall w \in W; \quad (9)$$

$$\sum_{k \in K} \sum_{w \in W} (x_{ik} \cdot y_{iw} \cdot z_{wk}) = p_i, \forall i \in N; \quad (10)$$

$$\sum_{k \in K} (k \cdot x_{ik}) \leq \sum_{k \in K} (k \cdot x_{jk}) + B \cdot (2 - y_{in} - y_{jn}), \quad \forall i, j \in N, \forall w \in W; \quad (11)$$

$$\sum_{k \in K} (k \cdot x_{jk}) \leq \sum_{k \in K} (k \cdot x_{ik}) + B \cdot (2 - y_{in} - y_{jn}), \quad \forall i, j \in N, \forall w \in W; \quad (12)$$

$$e_i \cdot x_{ik} \cdot y_{iw} = e_j \cdot x_{jk} \cdot y_{jv}, \quad \forall i, j \in N, \forall k \in K, \forall w \in W; \quad (13)$$

$$wc_w = \max_{i \in N} (c_i \cdot x_{ik} \cdot y_{iw} \cdot z_{wk}), \quad \forall k \in K, \forall w \in W; \quad (14)$$

$$ct_k = \max_{w \in W} (wc_w \cdot z_{wk}), \forall k \in K; \quad (15)$$

$$\sum_{w \in W} z_{wk} \leq mw, \forall k \in K. \quad (16)$$

其中:式(1)为综合优化目标 F ,由3个子优化目标 f_1 、 f_2 、 f_3 组成; α_1 、 α_2 、 α_3 分别为采用层次分析法^[13]求得的各子优化目标的权重系数,取值分别为(0.811 1, 0.486 7, 0.324 4)。同时为了减小子优化目标之间的数量级差异,本文根据问题规模设置了各子优化目标的标注化系数 g_1 、 g_2 、 g_3 ,使得综合优化结果中3个目标值更加平衡。式(2)中 w_1 和 w_2 分别为工作站数量和工人数量的权重系数,取值为0.8和0.2。式(3)表示工人负荷平滑度 f_2 ,其值越小,工人之间的负荷越均衡。式(4)表示工作站间混流产品的负载差异性平滑度 f_3 ,其中 $\sum_{i \in N} (x_{ik} \cdot dev_i)$ 用于描述工作站 k 上所有工序的总负载差异性指数, $\sum_{i \in N} dev_i / \|K\|$ 用于描述所有工序在各个工作站上的平均负载差异性指数,工作站间混流产品的负载差异性平滑度值越小,工作站间的负载越均衡。约束条件中:式(5)表示任意一道工序必须且只能分配到一个工作站上。式(6)表示任意一道工序不能在其紧前工序装配完成之前开始装配。式(7)表示一名工人不能同时加工两道工序。式(8)表示工序的完工时间 c_i 既不能早于工序的平均处理时间 \bar{t}_i ,也不能超过生产节拍 ct 。式(9)表示工人的总装配时间既不能早于分配给该工人的所有工序的平均处理时间之和,也不能超过工人的完工时间。式(10)~(13)为工序指派约束:式(10)表示指派给工序的工人类型和数量必须与工序所需工人类型和数量一致;式(11)和(12)表示指派给同一个工人的工序必须分配到同一个工作站;式(13)表示分配给同一个工人的两道工序所需工人类型必须相同。式(14)表示工人的完工时间等于其装配的所有工序中完工时间的最大值。式(15)表示工作站生产节拍等于该工作站内所有工人完工时间的最大值。式(16)表示工作站上的实际工人数量不得超过其所能容纳的最大工人数量。

2 基于DDQN的多人协同混流装配线平衡优化算法

2.1 马尔科夫决策模型

MCMALBP是一个离散序列性决策问题,即装配线每完成一道工序或产品的分配,都对应着当前时刻的生产状态,且 t 时刻的下一个生产状态 s_{t+1} 只与该时刻的生产状态 s_t 有关,而与 t 时刻之前的生产状态无关,满足马尔科夫性。因此,可以将 MCMALBP 的求解过程转化为马尔科夫决策过程,其主要包含状态空间、动作空间以及奖励函数。

2.1.1 MCMALBP的状态空间设计

考虑到MCMALBP是一个复杂的组合优化问题,且实际混流装配环境涉及到产品、工序、工作站、工人等多个对象,获取全部的生产信息是不合理且低效的.因此,根据MCMALBP生产状态的相关参数,结合工序分配过程中能够实时反映混流装配线生产状态的特征量以及相关文献^[14],设计MCMALBP的状态特征,如下列各式所示:

$$f_{b,1}(t) = \begin{cases} 0, & \sum_{w \in W} z_{wk_{\max}} = 0; \\ \frac{\sum_{w \in W} (z_{wk_{\max}} \cdot t_w)}{\sum_{w \in W} z_{wk_{\max}} \cdot ct}, & \text{otherwise.} \end{cases} \quad (17)$$

$$f_{b,2}(t) = \sum_{i \in N} (x_{ik_{\max}} \cdot dev_i). \quad (18)$$

$$f_{b,3}(t) = \|AO\|. \quad (19)$$

$$f_{b,4}(t) = \|CO\|/\|N\|. \quad (20)$$

$$f_{b,5}(t) = \sum_{i \in CO} \bar{t}_i / \sum_{i \in N} \bar{t}_i. \quad (21)$$

$$f_{b,6}(t) = \begin{cases} 0, & k_{\max} = 0; \\ \frac{\sum_{i \in CO} (\bar{t}_i \cdot p_i)}{\|W\| \cdot ct}, & \text{otherwise.} \end{cases} \quad (22)$$

$$f_{b,7}(t) = \begin{cases} 0, & k_{\max} = 0; \\ f_1, & \text{otherwise.} \end{cases} \quad (23)$$

$$f_{b,8}(t) = \begin{cases} 0, & k_{\max} = 0; \\ f_2, & \text{otherwise.} \end{cases} \quad (24)$$

$$f_{b,9}(t) = \begin{cases} 0, & k_{\max} = 0; \\ f_3, & \text{otherwise.} \end{cases} \quad (25)$$

其中: $f_{b,1}(t) \sim f_{b,3}(t)$ 为局部状态特征; $f_{b,4}(t) \sim f_{b,9}(t)$ 为全局状态特征; t 表示混流装配环境与平衡智能体交互的次数,且 $t = \{1, 2, \dots, \|N\|\}$, 例如 $t = 1$ 表示环境与智能体第1次交互,完成装配线上第1道工序的决策, $t = \|N\|$ 表示最后一道工序的决策. $f_{b,1}(t)$ 用于描述最大索引工作站上的工人空闲程度; $f_{b,2}(t)$ 用于描述最大索引工作站上所有工序在不同类型产品上的负载差异性指数之和; $f_{b,3}(t)$ 用于描述工序分配过程中的可选工序集合的大小; $f_{b,4}(t)$ 和 $f_{b,5}(t)$ 分别用于描述装配线在工序数量和处理时间上的完成度; $f_{b,6}(t)$ 用于描述装配线的线效率; $f_{b,7}(t) \sim f_{b,9}(t)$ 分别对应于式(2)~(4)中的3个子目标.

2.1.2 MCMALBP的动作空间设计

考虑到工序的装配过程具有连续性特点,即只有在当前工序装配完成后才能进行下一道工序的决策,因此,本文将每道工序的分配时刻定义为一个调度决策点,平衡智能体在每个调度决策点与混流装配环境进行交互,基于环境状态做出最佳的工序决策,从而得到各个工作站上的最优工序分配方案以及工人指派方案.同时,考虑到工序的装配顺序、处理时间、所需工人类型以及数量等属性对多人协同混流装配线的负载均衡性有着很大的影响,本文基于综合优先关系图中工序的处理时间权重、后续工序数量权重、综合位置权重、紧后工序数量、紧后工序处理时间以及工序的负载差异性指数、所需工人类型、所需工人数量等启发式信息,定义了10个复合动作策略 $Relu_{b,1} \sim Relu_{b,10}$ 来完成工序的决策.每种决策方式均能够在当前环境下确定下一道工序的分配.

2.1.3 MCMALBP的奖励函数设计

在混流装配线平衡问题的求解过程中,通常在新增工人或工作站时会直接影响工作站内工人的平均利用率,而装配方案的好坏往往体现在目标函数值的大小上,因此,本文基于MCMALBP的局部状态特征 $f_{b,1}(t)$ 以及优化目标 F 来设计奖励函数,即

$$R(t) = \begin{cases} -0.01, & t \neq \|N\| \text{ 且 } f_{b,1}(t) > f_{b,1}(t+1); \\ 0.01, & t \neq \|N\| \text{ 且 } f_{b,1}(t) \leq f_{b,1}(t+1); \\ -F, & t = \|N\|. \end{cases} \quad (26)$$

其中 $f_{b,1}(t)$ 和 $f_{b,1}(t+1)$ 分别为第 t 和 $t+1$ 道工序决策时最大索引工作站上的工人平均利用率.若 $f_{b,1}(t) > f_{b,1}(t+1)$,即当前动作决策执行后,使得最大索引工作站上的工人平均利用率下降,则奖励值为负,否则奖励值为正;同时,当所有工序决策完成之后,即 $t = \|N\|$ 时,根据状态特征 $f_{b,7}$ 、 $f_{b,8}$ 和 $f_{b,9}$ 计算目标函数值 F ,且目标函数值 F 越小,相应的奖励值越大.

2.2 算法设计

2.2.1 DDQN算法

本文采用 Double-DQN(DDQN) 算法^[15]. DDQN 算法主要有以下3个特点:

1) DDQN 算法采用预测网络与目标网络的双网络结构.两者具有相同的网络结构,但网络参数和更新方式有所不同.其中:预测网络的损失函数 $L_t(\theta_t)$ 为

$$L_t(\theta_t) = E[(r_t - Q(s_t, a_t; \theta_t) +$$

$$\gamma \cdot \hat{Q}(s_{t+1}, \arg \max_{a_{t+1} \in A} Q(s_{t+1}, a_{t+1}; \theta_t; \hat{\theta}_t))^2, \quad (27)$$

在每个训练步长都要对其参数 θ 进行更新;而目标网络的参数 $\hat{\theta}$ 则是每隔固定步长 c ,采用预测网络参数 θ 对其进行更新. DDQN 算法通过预测网络的动作价值函数 Q 来完成智能体的动作决策,而利用目标网络 \hat{Q} 来评价所选动作的优劣,从而有效避免模型更新时的过估计问题.

2) 训练过程中加入了经验回放池,用于存储训练过程中智能体与环境的交互数据. 从经验回放池中随机抽取小批量数据作为预测网络模型的训练样本,可以降低数据之间的关联程度,提高经验数据的使用效率,从而增加学习过程的平滑度.

3) 采用 ϵ -greed的动作决策,即

$$a_t = \begin{cases} \arg \max_{a \in A} Q, & \text{以 } 1 - \epsilon \text{ 的概率选取;} \\ \text{random}, & \text{以 } \epsilon \text{ 的概率选取.} \end{cases} \quad (28)$$

ϵ -greed 策略是一个探索与利用的过程,探索过程通常是随机的,其目的在于挖掘动作空间中能够获取到更大 Q 值的动作.

传统的 DDQN 算法通常采用随机的动作探索策略 (DDQN-random) 来完成动作选择,而本文采用一

种自适应的动作探索策略 (DDQN-adaptive),即

$$\epsilon = \epsilon_{\max} - (\epsilon_{\max} - \epsilon_{\min}) \cdot \min \left(1, \frac{2e_{\text{cur}}}{\text{EPI}} \right). \quad (29)$$

其中: ϵ_{\max} 为最大动作探索概率, ϵ_{\min} 为最小动作探索概率, e_{cur} 为模型当前的训练次数,EPI为模型的总训练次数.

2.2.2 基于 DDQN 强化学习的平衡优化过程

本文基于 DDQN 算法设计的混流装配线平衡优化框架如图 1 所示,包括混流装配环境和平衡智能体两部分. 其中:混流装配环境用于描述工序、工人和 workstation 等生产对象;平衡智能体观察混流装配环境中的状态特征,将当前环境状态输入到平衡智能体中,得到动作价值回报,随后根据 ϵ -greed 动作选择策略决策出动作并映射为相应的调度规则,再根据调度规则从混流装配环境的可选工序集中选择相应的工序,并结合 workstation 和工人的生产状态,将工序分配到合适的工作站上并指派相应的工人. 随后更新混流装配环境,此时平衡智能体将会收到环境的下一个状态特征以及当前动作策略对应的奖励值. 最后,采用式(27)更新神经网络模型的参数,并进行下一次工序分配的决策. 当完成所有工序的决策后,即可得到混流产品的工序分配和工人指派方案.

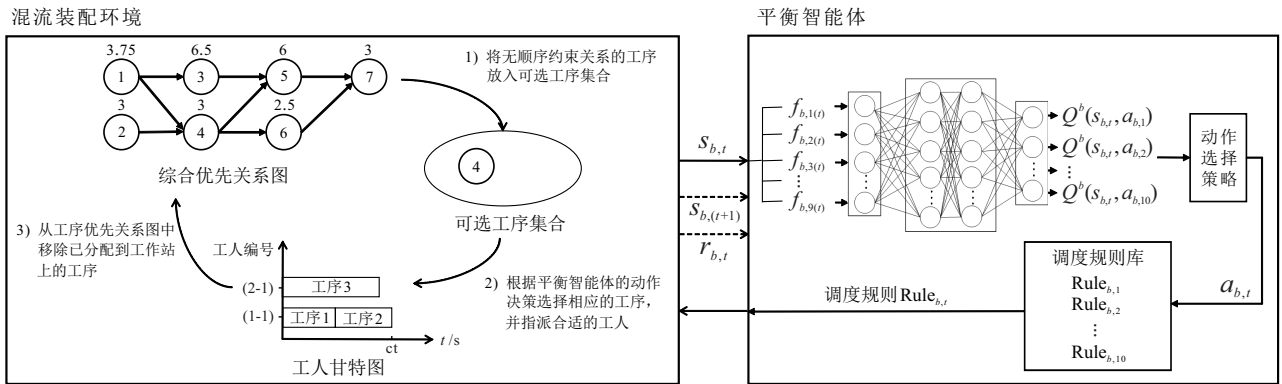


图 1 基于 DDQN 强化学习的平衡优化框架

2.2.3 考虑工人利用率的解码方法

平衡智能体每完成一次动作决策都对应着一道工序的选择,随后需要根据工序所需的工人类型和数量、workstation 上的时间以及空间容量等约束,将工序分配到合适的工作站上并为其指派相应的工人,从而得到工序的具体实施方案,即工序所分配到的 workstation 编号、工人编号集合、开始时间以及完工时间. 本文以 workstation 上的工人平均利用率为启发式信息,根据实际调度经验将工人平均利用率下限值 \bar{u}_{\min} 设置为 0.9,当智能体完成工序的决策后,优先考虑直接指派当前 workstation 上现有的空闲工人完成工序的装配,尽可能使当前 workstation 上的工人利用率达到最大. 若当前 work-

station 无法满足工序的装配工人需求,或新增工人后 workstation 的工人平均利用率小于下限值,则考虑新增一个 workstation 分配该工序.

3 实验设计与结果分析

3.1 算例生成

本文所研究的多人协同混流装配线平衡问题 (MCMALBP) 尚不存在标准算例数据,为了验证本文设计的 DDQN 算法的有效性,首先对混流装配线平衡问题 (MALBP) 的公开算例数据 (<https://assembly-line-balancing.de>) 进行求解. 在公开算例数据中,包含大中小 3 种规模的测试用例. 其中:小规模用例包含 19 道工序和 3 种产品,中规模用例包含 61 道工

序和4种产品,大规模用例包含111道工序和5种产品.但数据集中未给出生产节拍 ct 以及最小生产循环MPS.因此,本文采用文献[9]中给定工作站数量的20组不同规模的混流装配线标准测试用例进行实验,包含4组小规模测试用例,6组中规模测试用例,10组大规模测试用例.考虑到本文所研究问题的特点是给定生产节拍,因此根据公式

$$ct_{\min} = \frac{\sum_{m \in M} \sum_{i \in N} (d_m \cdot t_{mi})}{\sum_{m \in M} d_m \cdot \text{stations}}$$

计算最小生产节拍 ct_{\min} ,添加一定裕量后,得到各组测试用例的生产节拍 ct .20组MALBP标准测试用例的基本信息如表2所示,其中 g_1, g_2, g_3 为式(1)中综合优化目标函数的标准化系数.

表2 MALBP标准测试用例基本信息

No.	station	ct_{\min}	ct	MPS	g_1	g_2	g_3
1	3	2	2	1,1,1	1/3	1	10
2	3	1.883	2	3,2,1	1/3	1	10
3	4	1.5	1.6	1,1,1	1/4	1	10
4	4	1.412	1.6	3,2,1	1/4	1	10
5	5	17.58	18.36	1,1,1,1	1/5	1	10
6	5	18.287	18.36	1,3,4,5	1/5	1	1
7	7	12.557	13.18	1,1,1,1	1/7	1	10
8	7	13.062	13.18	1,3,4,5	1/7	1	1
9	10	8.79	9.35	1,1,1,1	1/10	1	5
10	10	9.144	9.35	1,3,4,5	1/10	1	1
11	9	16.192	16.4	1,1,1,1,1	1/9	1	1
12	9	16.104	16.4	1,2,4,5,8	1/9	1	2
13	9	16.273	16.4	5,3,2,1,1	1/9	1	1
14	9	16.165	16.4	1,4,8,3,1	1/9	1	1
15	13	11.21	11.4	1,1,1,1,1	1/13	1	1
16	13	11.149	11.4	1,2,4,5,8	1/13	1	2
17	13	11.266	11.4	5,3,2,1,1	1/13	1	1
18	13	11.191	11.4	1,4,8,3,1	1/13	1	1
19	15	9.715	9.9	1,1,1,1,1	1/15	1	1
20	15	9.663	9.9	1,2,4,5,8	1/15	1	2

为了验证本文所提出算法对于MCMALBP模型的有效性,在MALBP标准测试用例的基础上,生成带有工人数量 p_i 、工人类型 $\|E\|$ 以及工作站可容纳的最大工人数量 mw 等特点的MCMALBP测试用例.其中:4组小规模算例 p_i 的取值范围为 $1 \sim 2$, $mw = 3$, $\|E\| = 2$; 6组中规模算例 p_i 的取值范围为 $1 \sim 3$, $mw = 7$, $\|E\| = 3$; 10组大规模算例 p_i 的取值范围为 $1 \sim 5$, $mw = 10$, $\|E\| = 3$.根据测试用例的工序特征设置标准化系数 g_1, g_2, g_3 ,如表3所示.

3.2 算法参数设置

本文的实验环境为Windows10、i5-4460CPU、CPU频率3.2GHz、Python3.8,使用Pytorch框架和Python语言进行编程.本文中DDQN算法超参数设

表3 MCMALBP测试用例基本信息

No.	mw	$\ E\ $	ct	MPS	g_1	g_2	g_3
1	3	2	2	1,1,1	1/3	1	1
2	3	2	2	3,2,1	1/3	1	1
3	3	2	1.6	1,1,1	1/3	1	1
4	3	2	1.6	3,2,1	1/3	1	1
5	7	3	18.36	1,1,1,1	1/6	1/8	1/10
6	7	3	18.36	1,3,4,5	1/6	1/8	1/10
7	7	3	13.18	1,1,1,1	1/8	1/5	1/10
8	7	3	13.18	1,3,4,5	1/8	1/5	1/10
9	7	3	9.35	1,1,1,1	1/10	1/5	1/8
10	7	3	9.35	1,3,4,5	1/10	1/5	1/8
11	10	3	16.4	1,1,1,1,1	1/20	1/10	1
12	10	3	16.4	1,2,4,5,8	1/20	1/10	1
13	10	3	16.4	5,3,2,1,1	1/20	1/10	1
14	10	3	16.4	1,4,8,3,1	1/20	1/10	1
15	10	3	11.4	1,1,1,1,1	1/20	1/10	1
16	10	3	11.4	1,2,4,5,8	1/20	1/10	1
17	10	3	11.4	5,3,2,1,1	1/20	1/10	1
18	10	3	11.4	1,4,8,3,1	1/20	1/10	1
19	10	3	9.9	1,1,1,1,1	1/20	1/10	1
20	10	3	9.9	1,2,4,5,8	1/20	1/10	1

置主要是在参考了相关文献^[6]的参数设置后完成的,其中预测网络与训练网络的输入层、输出层、隐藏层节点个数分别为9、10、16,隐藏层层数为2,训练次数 $EPI = 1000$,经验池容量1000,激活函数采用Relu,优化器采用Adam,奖励函数的折扣因子 γ 为0.95,学习率为0.005,动作的最大和最小探索概率 $\epsilon_{\max}, \epsilon_{\min}$ 分别为0.9和0.1,网络更新步长 c 为20.

本文选取多人共站装配线平衡问题中表现较好的模拟退火(simulated annealing, SA)算法^[17]和改进离散水波优化(discrete water wave optimization, DWWO)算法^[18]作为对比算法,两种算法均采用文献[18]中基于拓扑排序的编码方式生成初始种群,适应度函数根据综合优化目标设置为 $100/F$.DWWO和SA的参数分别按相关文献进行设置,两种算法的具体参数值如表4所示.

表4 DWWO算法与SA算法参数

算法	参数名称
DWWO	种群大小:200 迭代次数:500 初始波长:1 扰动概率:0.25 最大高度:16 波长碎浪系数:0.2 波长衰减系数:1.001
	种群大小:200 中小规模问题终止温度:0.001 中小规模问题初始温度:1000 中小规模问题退火速度:0.97 大规模问题的初始温度:2000 大规模问题的终止温度:10 大规模问题的退火速度:0.997

3.3 算法有效性验证

3.3.1 自适应探索概率的动作选择策略有效性验证

为了验证本文采用的自适应探索概率的动作选择策略有效性,本文从MCMALBP测试用例中选择

小规模测试用例2、中规模测试用例6和大规模测试用例12,对比动作探索下降速率为0.01的随机探索概率的动作选择策略(DDQN-random)与自适应探索概率的动作选择策略(DDQN-adaptive)在模型训练过程的奖励函数收敛曲线,如图2~图4所示。观察图2~图4可知,DDQN-random策略和DDQN-adaptive策略在模型训练过程中的最大奖励函数值大致相同,

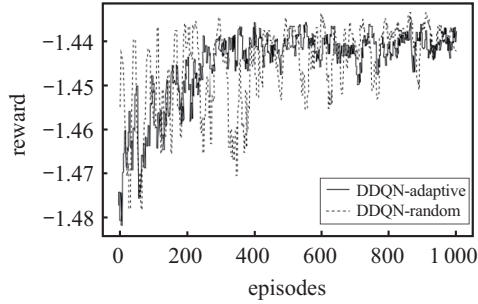


图2 MCMALBP测试用例2的奖励函数收敛曲线

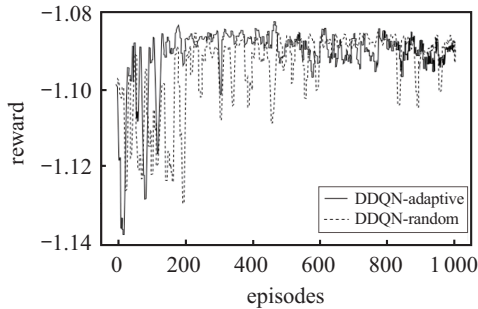


图3 MCMALBP测试用例6的奖励函数收敛曲线

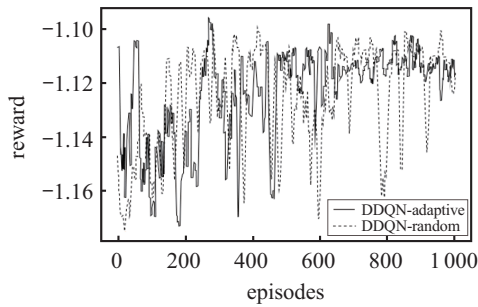


图4 MCMALBP测试用例12的奖励函数收敛曲线

但在DDQN-adaptive策略下,奖励函数值的整体波动性更小,说明该策略在算法稳定性上表现更优。故后续实验中,DDQN算法均采用DDQN-adaptive策略进行动作选择。

3.3.2 动作策略有效性实验

为了验证所选动作策略的有效性,考虑单独移除每个动作后,保留剩下动作进行实验,即表5中的 $\overline{Relu}_{b,1}$ 列表示不采用动作 $Relu_{b,1}$ 的实验结果,以此类推。为确保实验的准确性,本文从MCMALBP算例中选择部分产品需求不相等的算例,并对每个算例分别重复运行20次,统计20次运行结果的平均值并展示于表5中。从表5中可以看出,不移除任何动作的DDQN算法具有最优的性能,而其他移除了某一动作策略的实验结果均有一定程度的下降,从而表明了本文选择的动作特征的有效性。

3.4 MALBP实验

为了测试所提出算法的有效性,本文采用DDQN算法、DWWO算法以及SA算法对表2中的MALBP标准测试用例进行求解,3种算法在每个测试用例上独立运行20次,实验结果如表6所示。其中: F_{best} 、 F_{avg} 、 F_{std} 分别表示算法重复运行20次求得的综合优化目标最优值、平均值和标准差,每个算例的最好结果均用粗体标出。

由表6可知:在20组不同规模的MALBP标准测试用例上,DDQN算法能够在14组测试用例(占70%)中求得综合优化目标的最优值,相对于DWWO(8组测试用例,占40%)和SA(6组测试用例,占30%),DDQN算法具有更好的整体寻优性能;同时,DDQN算法能够在16组测试用例(占80%)中求得优于其他两种算法的综合优化目标的平均值,在15组(占75%)测试用例中求得优于其他两种算法的标准差,说明DDQN算法具有更好的求解精度和稳定性。

表5 动作策略有效性测试的实验结果

No.	DDQN	$\overline{Relu}_{b,1}$	$\overline{Relu}_{b,2}$	$\overline{Relu}_{b,3}$	$\overline{Relu}_{b,4}$	$\overline{Relu}_{b,5}$	$\overline{Relu}_{b,6}$	$\overline{Relu}_{b,7}$	$\overline{Relu}_{b,8}$	$\overline{Relu}_{b,9}$	$\overline{Relu}_{b,10}$
2	1.1137	1.1204	1.1187	1.1168	1.1165	1.1181	1.1179	1.1161	1.1169	1.1188	1.1196
4	1.3774	1.3831	1.3833	1.3875	1.3809	1.3910	1.3939	1.3900	1.3891	1.3908	1.3863
6	1.4552	1.4758	1.4718	1.4710	1.4697	1.4682	1.4622	1.4633	1.4621	1.4651	1.4633
8	1.3354	1.3486	1.3431	1.3510	1.3482	1.3422	1.3439	1.3417	1.3431	1.3458	1.3472
10	1.1393	1.2095	1.2353	1.2390	1.2082	1.2558	1.2628	1.2635	1.2504	1.2626	1.2605
12	1.1463	1.1752	1.2066	1.2065	1.1943	1.2085	1.2054	1.2030	1.2031	1.2055	1.2051
16	1.1470	1.1574	1.1527	1.1540	1.1549	1.1631	1.1640	1.1637	1.1615	1.1606	1.1628
20	1.1618	1.1743	1.1693	1.1668	1.1683	1.1721	1.1692	1.1662	1.1691	1.7103	1.6813

表6 3种算法求解MALBP测试用例的实验结果

No.	$\ N\ $	DDQN			DWWO			SA		
		F_{best}	F_{avg}	F_{std}	F_{best}	F_{avg}	F_{std}	F_{best}	F_{avg}	F_{std}
1	19	0.845 2	0.858 1	0.018 7	0.845 2	0.867 1	0.021 7	0.845 2	0.869 9	0.022 3
2	19	0.894 7	0.899 7	0.006 3	0.894 7	0.899 7	0.006 3	0.894 7	0.904 8	0.008 9
3	19	0.890 9	0.893 1	0.002 0	0.890 9	0.898 7	0.012 3	0.890 9	0.908 2	0.019 1
4	19	0.944 3	0.966 9	0.018 8	0.944 3	0.985 5	0.022 0	0.944 3	0.996 7	0.023 2
5	61	1.273 2	1.347 3	0.029 6	1.250 0	1.306 1	0.046 3	1.304 1	1.516 7	0.131 6
6	61	0.915 3	1.172 4	0.887 0	0.917 1	1.161 0	0.911 1	0.958 2	1.653 3	1.497 5
7	61	1.197 5	1.300 9	0.055 0	1.208 9	1.312 5	0.095 2	1.193 7	1.586 1	0.309 1
8	61	0.947 6	1.021 9	0.040 3	0.936 9	1.035 3	0.064 3	0.948 4	1.049 0	0.087 6
9	61	1.195 7	1.226 1	0.015 8	1.205 1	1.288 5	0.061 3	1.257 9	1.388 8	0.102 0
10	61	0.959 0	1.043 2	0.120 1	0.961 0	1.027 0	0.120 6	0.966 7	1.159 4	0.178 5
11	111	0.985 8	1.096 0	0.271 7	0.987 2	1.151 5	0.323 4	0.993 8	1.248 4	0.366 4
12	111	1.023 8	1.039 7	0.008 9	1.024 4	1.042 1	0.008 4	1.026 2	1.049 4	0.027 1
13	111	0.978 0	1.131 9	0.312 5	0.977 6	1.220 0	0.377 2	0.974 2	1.135 0	0.318 9
14	111	1.003 7	1.018 2	0.008 0	1.004 3	1.020 7	0.007 3	1.005 3	1.025 8	0.010 7
15	111	0.978 1	1.032 1	0.138 7	0.977 2	1.122 6	0.210 8	0.982 0	1.193 0	0.217 7
16	111	1.004 2	1.010 9	0.004 1	1.005 7	1.012 7	0.003 3	1.006 3	1.011 2	0.005 3
17	111	0.972 0	1.028 7	0.135 3	0.976 6	1.098 0	0.197 2	0.973 3	1.046 9	0.159 9
18	111	0.986 5	1.061 2	0.129 7	0.991 1	1.078 8	0.157 0	0.989 8	1.061 4	0.162 8
19	111	0.973 2	1.001 4	0.073 8	0.974 6	1.008 4	0.072 7	0.975 9	1.016 3	0.102 7
20	111	1.004 7	1.077 2	0.135 7	1.003 2	1.048 2	0.100 3	1.005 0	1.101 5	0.169 0

表7 3种算法求解MCMALBP测试用例的实验结果

No.	$\ N\ $	DDQN			DWWO			SA		
		F_{best}	F_{avg}	F_{std}	F_{best}	F_{avg}	F_{std}	F_{best}	F_{avg}	F_{std}
1	19	1.230 3	1.268 3	0.016 2	1.240 2	1.274 0	0.020 5	1.250 7	1.272 6	0.028 8
2	19	1.080 7	1.113 7	0.011 8	1.075 0	1.136 8	0.031 8	1.068 5	1.100 1	0.017 9
3	19	1.477 2	1.483 3	0.008 6	1.488 9	1.513 8	0.017 1	1.490 8	1.514 9	0.012 3
4	19	1.295 6	1.377 4	0.073 6	1.297 2	1.425 9	0.108 4	1.297 4	1.411 1	0.100 3
5	61	1.427 9	1.443 6	0.010 3	1.443 2	1.466 1	0.016 0	1.497 7	1.531 2	0.027 2
6	61	1.433 5	1.455 2	0.016 0	1.407 0	1.450 4	0.016 6	1.414 3	1.460 4	0.018 6
7	61	1.306 6	1.353 9	0.032 4	1.321 8	1.388 1	0.030 0	1.334 6	1.402 9	0.029 5
8	61	1.297 0	1.335 4	0.019 1	1.298 5	1.367 1	0.037 9	1.329 9	1.383 6	0.025 9
9	61	1.231 7	1.258 1	0.015 4	1.187 1	1.253 0	0.045 2	1.215 5	1.264 5	0.029 1
10	61	1.117 7	1.139 3	0.025 6	1.121 5	1.173 5	0.038 0	1.124 7	1.174 9	0.049 1
11	111	1.058 8	1.194 1	0.046 6	1.091 4	1.197 4	0.033 5	1.109 1	1.209 0	0.047 3
12	111	1.110 9	1.146 3	0.027 6	1.148 2	1.183 4	0.024 0	1.162 9	1.178 7	0.021 1
13	111	1.173 7	1.223 9	0.022 7	1.223 1	1.288 4	0.023 1	1.216 3	1.291 0	0.060 6
14	111	1.181 7	1.198 4	0.012 2	1.192 3	1.223 1	0.022 7	1.193 4	1.257 5	0.060 4
15	111	1.176 1	1.204 0	0.022 7	1.163 9	1.239 3	0.063 6	1.182 2	1.266 9	0.054 7
16	111	1.088 8	1.147 0	0.035 3	1.094 2	1.150 9	0.034 9	1.119 5	1.207 3	0.055 4
17	111	1.153 6	1.183 5	0.022 5	1.152 6	1.193 9	0.028 8	1.161 7	1.234 2	0.050 9
18	111	1.128 4	1.160 7	0.020 7	1.131 5	1.210 8	0.045 5	1.129 3	1.201 2	0.043 5
19	111	1.157 6	1.198 5	0.026 7	1.166 9	1.222 5	0.025 6	1.198 5	1.257 8	0.035 1
20	111	1.145 7	1.161 8	0.021 1	1.141 4	1.168 2	0.014 3	1.179 5	1.179 5	0.029 3

3.5 MCMALBP 实验

为了验证本文建立的 MCMALBP 优化模型的有效性以及 DDQN 算法对于该模型的求解性能, 分别采用 DDQN 算法、DWWO 算法以及 SA 算法对表 3 中的测试用例进行求解, 3 种算法在每个测试用例上独立运行 20 次, 表 7 给出了各算法对应的综合优化目标的最优值、平均值和标准差, 并用粗体标明各个指标的最好结果。

由表 7 可知: 在 20 组不同规模的 MCMALBP 测试用例上, DDQN 算法能够在 14 组测试用例 (占 70%) 中求得最优的综合优化目标值; 在 16 组中、大规模测试用例上, DDQN 算法在 68.5% 的测试用例中求得最优解; 而 DWWO 算法和 SA 算法分别在 31.25% 和 0% 的测试用例中求得最优解。同时, 在实验过程中发现: 针对小规模算例, 3 种算法运行时间均在 50 s 左右, 无明显差异; 针对中规模算例, DDQN 算法的运行时间约为 300 s, DWWO 算法运行时间约为 1 000 s, SA 算法运行时间约为 800 s; 针对大规模算例, DDQN 算法运行时间约为 700 s, 而 DWWO 算法运行时间约为 2 900 s, SA 算法运行时间约为 2 500 s。由运行时间可以看出 DDQN 算法在中规模 and 大规模算例上搜索效率高。

为了更好地说明算法在不同规模测试用例上的整体性能, 计算表 7 中 3 种算法在 20 组测试用例上所得 3 个子优化目标和线效率 LE 的平均值, 统计结果如表 8 所示。其中线效率 LE 由下式计算:

$$LE = \frac{\sum_{m \in M} \left(d_m \cdot \sum_{i \in N} (t_{mi} \cdot p_i) \right)}{\|W\| \cdot ct} \cdot 100\% \quad (30)$$

由式 (30) 可知, 在给定生产节拍下, 装配线上的工人数量越少, 分配给每个工人的工序总处理时间越多, 工人的时间利用率越高, 装配线线效率也越高。

表 8 3 种算法的整体优化性能指标对比

算法	$f_{1,avg}$	$f_{2,avg}$	$f_{3,avg}$	LE _{avg}
DDQN	13.5485	3.9363	2.1328	0.7062
DWWO	13.7840	4.0609	2.1163	0.6872
SA	14.0130	4.1355	2.2303	0.6853

由表 8 可知, 相比于 DWWO 和 SA 算法, DDQN 算法在 20 组测试用例上的平均工作站数量和工人数量的综合指数、平均工人负载平滑度更小, 平均线效率更大, 说明 DDQN 算法在优化这 3 个指标上的效果最好, 而 DWWO 算法对于平均产品负载差异性平滑度的优化效果最好。

3.6 车辆混流装配线实例优化

为了适应生产特点和管理需求, Z 公司在车身装配过程中, 将所有工人按照其掌握的技能划分为不同的班组进行管理, 包含钳工班组、电工班组和管道工班组, 根据订单需求量完成工人当天的排班计划。对 Z 公司进行调研后得知, 目前 Z 公司订单需求量较多的是 A 系列车辆, 包含 4 种车型, 各产品需求比例为 5:2:3:2, 车身装配过程共包含 30 道工序, 工序所需工人数量 p_i 的取值范围为 1~3。同时, 考虑到公司的生产管理制度, 生产节拍 $ct = 8h$, 工作站所能容纳的最大工人数量 $mw = 5$, 工人种类 $\|E\| = 3$ 。当接收到车辆订单后, 由生产部门基于启发式规则中的阶位法得到订单的平衡方案, 如表 9 所示。

表 9 优化前后的平衡方案性能指标对比

平衡方案	$\ K\ $	$\ W\ $	WSI	PSI	LE
优化前	7	28	1.6771	1.6993	0.8443
优化后	6	26	0.9720	0.9971	0.9093

采用本文提出的 DDQN 算法对同一订单进行优化后, 新的平衡方案如表 9 所示。相较于优化前的平衡方案, 优化后的平衡方案的工作站数量减少了 14%, 工人数量减少了 7%, 说明优化后方案在订单装配过程中的线长更短, 所需的工人成本也更低。同时, 优化后方案所对应的工人负载平滑度降低了 42.04%, 产品负载差异性平滑度降低了 41.32%, 线效率提高了 7.7%, 说明采用 DDQN 算法进行平衡优化后, 整体提升了混流装配线的负载均衡性和生产效率。

4 结 论

本文针对大型设备混流装配过程中的多人协同、多工种的情况, 提出了多人协同混流装配线平衡问题 (MCMALBP) 的数学模型。针对 MCMALBP 的特点, 结合产品、工序、工作站、工人等生产对象的特征, 基于综合位置权重、紧后工序数量等启发式规则, 建立了 MCMALBP 的马尔科夫决策模型, 并设计了基于 DDQN 的平衡优化算法以及考虑工人利用率的工序解码方法。在混流装配线标准测试用例和多人协同混流装配线测试用例上验证了模型的有效性和算法的优越性。基于 Z 公司车辆混流装配线的订单数据以及生产现状, 采用本文提出的 DDQN 算法进行平衡优化, 优化后所得的平衡方案各指标均得到了优化, 从而验证了本文模型的有效性与方法的实用性。

目前, 本文研究未考虑多人协同混流装配环境中的不确定性因素, 例如市场需求波动引起的订单变

化,装配过程中的工序处理时间和工人配合熟练度变化等.后续研究中可以考虑加入对不确定性因素的研究,以提高模型和算法的鲁棒性.

参考文献(References)

- [1] Kilbridge M D, Wester L. A heuristic method of assembly line balancing[J]. *Journal of Industrial Engineering*, 1961, 12(4): 292-298.
- [2] Sewell E C, Jacobson S H. A branch, bound, and remember algorithm for the simple assembly line balancing problem[J]. *INFORMS Journal on Computing*, 2012, 24(3): 433-442.
- [3] Jiao Y L, Cao N, Li J, et al. Balancing a U-shaped assembly line with a heuristic algorithm based on a comprehensive rank value[J]. *Sustainability*, 2022, 14(2): 775.
- [4] 刘俨后, 麻娟, 左敦稳, 等. 基于改进烟花算法的随机装配线混流调度[J]. *控制与决策*, 2019, 34(4): 751-758.
(Liu Y H, Ma J, Zuo D W, et al. Mixed-model sequencing of stochastic assembly line based on improved fireworks algorithm[J]. *Control and Decision*, 2019, 34(4): 751-758.)
- [5] Zhang B K, Xu L Y, Zhang J. A multi-objective cellular genetic algorithm for energy-oriented balancing and sequencing problem of mixed-model assembly line[J]. *Journal of Cleaner Production*, 2020, 244: 118845.
- [6] 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[J]. *控制与决策*, 2021, 36(11): 2609-2617.
(Wang L, Pan Z X. Scheduling optimization for flow-shop based on deep reinforcement learning and iterative greedy method[J]. *Control and Decision*, 2021, 36(11): 2609-2617.)
- [7] 孙爱红, 雷琦, 宋豫川, 等. 基于深度强化学习求解作业车间机器与AGV联合调度问题[J]. *控制与决策*, 2024, 39(1): 253-262.
(Sun A H, Lei Q, Song Y C, et al. Deep reinforcement learning for solving the joint scheduling problem of machines and AGVs in job shop[J]. *Control and Decision*, 2024, 39(1): 253-262.)
- [8] Luo S. Dynamic scheduling for flexible job shop with new job insertions by deep reinforcement learning[J]. *Applied Soft Computing*, 2020, 91: 106208.
- [9] Wu Y X, Song W, Cao Z G, et al. Learning improvement heuristics for solving routing problems[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2022, 33(9): 5057-5069.
- [10] 王思鹏, 杜昌平, 郑耀. 基于强化学习的扑翼飞行器路径规划算法[J]. *控制与决策*, 2022, 37(4): 851-860.
(Wang S P, Du C P, Zheng Y. Local planner for flapping wing micro aerial vehicle based on deep reinforcement learning[J]. *Control and Decision*, 2022, 37(4): 851-860.)
- [11] 赵铭慧, 张雪波, 郭宪, 等. 基于深度强化学习的双向装配序列规划[J]. *控制理论与应用*, 2021, 38(12): 1901-1910.
(Zhao M H, Zhang X B, Guo X, et al. Assembly sequence planning based on deep reinforcement learning[J]. *Control Theory & Applications*, 2021, 38(12): 1901-1910.)
- [12] Brammer J, Lutz B, Neumann D. Solving the mixed model sequencing problem with reinforcement learning and metaheuristics[J]. *Computers & Industrial Engineering*, 2021, 162: 107704.
- [13] Dos Santos P H, Neves S M, Sant' Anna D O, et al. The analytic hierarchy process supporting decision making for sustainable development: An overview of applications[J]. *Journal of Cleaner Production*, 2019, 212: 119-138.
- [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] van Hasselt H, Guez A, Silver D. Deep reinforcement learning with double Q-learning[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2016, 30(1): 2094-2100.
- [16] Hessel M, Modayil J, van Hasselt H, et al. Rainbow: Combining improvements in deep reinforcement learning[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2018, 32(1): 3215-3222.
- [17] Roshani A, Nezami F G. Mixed-model multi-manned assembly line balancing problem: A mathematical model and a simulated annealing approach[J]. *Assembly Automation*, 2017, 37(1): 34-50.
- [18] 张梅, 傅艳霞, 朱金辉, 等. 基于改进水波算法的复杂多人共站装配线平衡研究[J]. *计算机集成制造系统*, <https://kns.cnki.net/kcms/detail/11.5946.TP.20220119.1755.030.html>.
(Zhang M, Fu Y X, Zhu J H, et al. Complex multi-manned assembly line balancing using improved water wave algorithm[J]. *Computer Integrated Manufacturing Systems*, <https://kns.cnki.net/kcms/detail/11.5946.TP.20220119.1755.030.html>.)

作者简介

张梅(1977—),女,副教授,博士,从事智能优化与调度、智能算法与仿真等研究, E-mail: zhangmei@scut.edu.cn;

田镇遇(1997—),男,硕士生,从事智能优化与调度的研究, E-mail: autianzy@mail.scut.edu.cn;

朱金辉(1977—),男,副教授,博士,从事机器人软件、人工智能等研究, E-mail: csjhzhu@scut.edu.cn;

傅艳霞(1999—),女,硕士生,从事智能优化与调度的研究, E-mail: 1191584905@qq.com.