DOI:10.3969/j.issn.1003-5060.2025.07.003

基于DRL的大规模定制装配车间调度研究

屈新怀， 张慧慧， 丁必荣， 孟冠军
(合肥工业大学 机械工程学院,安徽 合肥 230009)

摘 要:针对大规模定制装配车间中订单的随机性和偶然性问题,文章提出一种基于深度强化学习(deep reinforcement learning,DRL)的大规模定制装配车间作业调度优化方法。建立以最小化产品组件更换次数和最小化订单提前/拖期惩罚为目标的大规模定制装配车间作业调度优化模型,基于调度模型建立马尔科夫决策过程,合理定义状态、动作和奖励函数;将调度模型优化问题与DRL方法相结合,并采用改进的D3QN算法进行模型求解;最后进行仿真实验验证。结果表明,文章所提方法能有效减少产品组件更换次数和降低订单提前/拖期惩罚。

关键词:大规模定制;装配车间;深度强化学习(DRL);车间作业调度;调度优化模型
中图分类号:TP301.6;TP391.9 **文献标志码**:A **文章编号**:1003-5060(2025)07-0878-06

Research on assembly shop scheduling for mass customization based on DRL

QU Xinhuai， ZHANG Huihui， DING Birong， MENG Guanjun
(School of Mechanical Engineering, Hefei University of Technology, Hefei 230009, China)

Abstract: Aiming at the randomness and contingency of orders in the assembly shop for mass customization, this paper proposes an assembly job shop scheduling optimization method based on deep reinforcement learning(DRL). Firstly, an assembly job shop scheduling optimization model is established to minimize the number of product component replacements and the penalties for order earliness or tardiness. Then, based on the scheduling model, a Markov decision process is established, and the state, action and reward functions are reasonably defined. The optimization problem of the scheduling model is solved by connecting it with the DRL method, and an improved D3QN algorithm is selected to solve the model. Finally, the simulation experiment is conducted, and the results show that the method effectively reduces the number of product component replacements and penalties for early or delayed orders.

Key words: mass customization; assembly shop; deep reinforcement learning(DRL); job shop scheduling; scheduling optimization model

0 引 言

在当前市场,客户需求日益多样化,订单品种日益繁多,大规模定制的生产方式结合了大规模生产效率和个性化定制特点,正日益成为企业首选的生产模式之一。在大规模生产过程中,部分车间为提高生产效率采用模块化设计[1]。将产品分解为模块,并根据模块特征分为通用件和个性件,在订单到达前按库存批量生产模块,在订单到达后按订单装配模块[2]。因此,装配车间作业调度问题的研究也日益受到重视。

在求解装配车间作业调度问题时,部分学者从不同角度进行优化,如减少能耗[3-4]、降低生产成本和缓冲区库存[5]、最小化工件完成时间[6]等。

收稿日期:2023-07-13;**修回日期**:2023-08-28
基金项目:国家重点研发计划资助项目(2019YFB1705303)
作者简介:屈新怀(1971—),男,安徽金寨人,博士,合肥工业大学副教授,硕士生导师.

许多学者采用元启发式算法求解装配车间作业调度问题，主要包括蚁群算法[7-8]、粒子群算法[9]、遗传算法[10]、模拟退火算法[11]等。这些算法通过随机搜索和局部优化寻找问题的近似最优解，并能根据问题的特性和约束条件设计不同的启发式规则和策略，适应不同的问题和场景。

大规模定制的订单具有加工时间和产品数量的不确定性以及产品组件的随机性。在装配过程中，每次更换产品组件类型都会增加个性件与车间的匹配成本以及库存取出成本；同时，顾客订单投入生产后必须在交货期内完成装配，以提高顾客满意度。研究表明，减少模块化组件更换次数不仅能节约成本，还能提高效率[12]，但目前关于这方面的装配车间作业调度研究较少。为此，本文以最小化产品组件更换次数和最小化订单提前/拖期惩罚为目标，建立基于深度强化学习(deep reinforcement learning，DRL)的装配车间作业调度模型。

1　问题描述与建模

1.1　问题描述

在面向订单装配的车间中，企业需要提供设备进行产品组装。考虑到订单数量庞大，部分企业会设置多个组装设备(如整车车间)。

由于订单数量多且组件品种多，设产品的订单组件中包含 i 种个性组件，每个个性组件有 j 种具体分类，其他组件为通用件；在实际装配时订单需要分配给 m 个设备，装配完成后再分配下一个订单。若在 1 个时间窗车间同时接收到 x 个订单，则第 x 个订单的交货期设为 d_x，订单中的产品组件记为 p_{ij}，第 x 个订单的产品数量为 n_x，第 x 个订单的总装配时间为 o_x。具体参数设置见表 1所列。

表 1　参数设置

符号	定义	内容
X	订单集合	$\{1,2,\cdots,x\}$
P	产品的组件集合	$\{p_{11},p_{12},\cdots,p_{1j};p_{21},p_{22},\cdots,p_{2j};\cdots;p_{i1},p_{i2},\cdots,p_{ij}\}$
N	订单的产品数量集合	$\{n_1,n_2,\cdots,n_x\}$
M	装配设备集合	$\{1,2,\cdots,m\}$
D	订单交付期集合	$\{d_1,d_2,\cdots,d_x\}$
T	订单的装配完成时间集合	$\{t_1,t_2,\cdots,t_x\}$
O	订单的装配时间集合	$\{o_1,o_2,\cdots,o_x\}$

根据大规模定制装配车间作业调度的特点，建立模型时做出以下假设条件：

1) 每批订单的产品组件种类相同；

2) 所有产品的总装配时间相同，组件类型不同不影响装配时间，每个产品组装的全过程为装配时间，订单的装配时间根据订单数量会有所不同；

3) 每个订单是相互独立的，安排每个订单的设备不影响其他订单分配；

4) 每个设备在零时刻均处于空闲状态；

5) 每个设备在加工一批订单时不再分配其他订单；

6) 各设备组装过程中不可中断；

7) 一批订单只能安排在一个设备上装配；

8) 本模型仅考虑装配车间个性件的组装环节，其他环节不考虑。

1.2　数学建模

基于模型假设，建立最小化产品个性件更换次数和最小化订单提前/拖期惩罚的装配车间作业调度模型，将最小化产品个性件更换次数转换成最大化个性件的相似度。

1) 总目标函数。x 个订单装配完成后的总目标函数描述为：

$$F=\min\sum_{\omega=0}^{x}\left(\frac{\gamma}{R_\omega+1}+\delta f_\omega\right),$$
$$\gamma+\delta=1,\quad 0\leqslant\delta\leqslant1,\quad 0\leqslant\gamma\leqslant1 \tag{1}$$

其中：R_ω 为产品个性组件的相似度；f_ω 为每个订单的提前/拖期惩罚；γ、δ 为权重参数；ω 为订单编号。

2) 提前/拖期惩罚函数。为了减少订单 ω 的装配完成时间超过交货期限而导致库存积压的损失，以及为了防止装配完成时间早于交货期限而导致客户满意度降低的损失，可以设置提前交货和拖期的惩罚函数。惩罚函数的数学模型可表示为：

$$f_\omega=\alpha\max(d_\omega-t_\omega,0)+\beta\max(t_\omega-d_\omega,0),$$
$$\alpha+\beta=1,\quad 0\leqslant\alpha\leqslant1,\quad 0\leqslant\beta\leqslant1 \tag{2}$$

其中：t_ω 为订单 ω 的加工结束时间；d_ω 为订单 ω 的交货期限；α 为提前惩罚因子；β 为拖期惩罚因子。

3) 个性组件的相似度。一个产品有 i 种个性件，当前设备装配的上一个订单的组件类型与订单 ω 的组件类型的相似度可表示为：

$$R_\omega=\sum_{a=1}^{i}\lambda_a s_a,\ \forall a \tag{3}$$

其中：λ_a 为组件 a 的相似度权重；s_a 表示组件 a

的相似度。

s_a 的数学模型表示为:

$$s_a=\begin{cases}0, & p_a^{\omega}\neq p_a^{\omega_1},\forall\omega;\\ 1, & p_a^{\omega}=p_a^{\omega_1},\forall\omega\end{cases} \tag{4}$$

其中:p_a^{ω}为订单ω内部个性组件a的种类;$p_a^{\omega_1}$为设备上一个订单ω_1内部个性组件a的种类。

4) 约束条件。模型需要满足的约束条件定义如下。

当前设备上一个订单的装配结束时间等于下一个订单的装配开始时间的数学模型描述为:

$$t_{\omega}=t_{\omega-1}+o_{\omega-1},\quad \forall\omega \tag{5}$$

其中:t_{ω}为订单ω的装配起始时间;$t_{\omega-1}$为当前设备上一个订单的装配起始时间;$o_{\omega-1}$为当前设备上一个订单的装配时间。

设备第1个订单的装配起始时间为0,具体描述为:

$$t_1=0 \tag{6}$$

x个订单全部完成装配描述为:

$$\sum_{\omega=1}^{x}U_{\omega}=0 \tag{7}$$

其中,U_{ω}为订单ω的调度状态,即

$$U_{\omega}=\begin{cases}0, & \text{订单}\omega\text{被分配};\\ 1, & \text{订单}\omega\text{未被分配}\end{cases} \tag{8}$$

2 算法设计

2.1 算法介绍

DRL是机器学习的一种特殊方法,它独立于监督学习和非监督学习,不需要经验数据支持,通过自己不断探索累积经验数据,学习状态与动作之间的关系,从而做出更优的决策。智能体(agent)通过观察环境的状态(state),选择一个动作(action)来影响环境,并根据环境返回的奖励信号(reward)来评估其行为的好坏。智能体的目标是找到一个最优策略,即在给定状态下选择最佳动作的策略,以最大化累积奖励(return)。

一些学者采用DRL求解装配车间调度问题,以提升算法学习和搜索效率[13-15]。在DRL的方法中,深度Q网络(deep Q-network,DQN)算法通过将深度学习引入强化学习,能够有效解决高维数据的抽象表征问题。该算法利用神经网络逼近Q值函数,从而获得特定状态下选择不同动作所对应的累计奖励。文献[16-17]采用改进的DQN算法求解航空发动机装配调度问题,均展现出良好的性能。

D3QN(dueling double deep Q-network)是DQN的改进算法,它结合了Dueling DQN中的分支结构和Double DQN中的双Q网络。其中:① 分支深度Q网络(Dueling DQN)将每个动作的Q值分解成状态值(value)和动作优势值(advantage)来提高决策能力;② 双层深度Q网络(Double DQN)与DQN不同,采用2个神经网络,即评估网络和目标网络。评估网络用于评估最优动作价值对应的动作,它以接收到的当前状态作为输入,并输出每个动作的Q值估计,评估网络的参数在每次状态更新时进行更新,使其能够根据当前策略选择最优动作,并通过减少过高估计Q值的影响来提高训练的稳定性和收敛性;目标网络与DQN的神经网络作用类似,用于计算所选择动作的Q值,目标网络的参数通常是通过定期拷贝评估网络的参数而得到的,并在一段时间内保持不变,从而降低训练的波动。本文选用D3QN算法求解大规模定制的装配车间作业调度问题。

2.2 算法实现

DRL利用马尔可夫决策过程(Markov decision process,MDP)的无记忆性观察当前状态和采取动作来预测下一个状态的转移概率,并根据奖励信号学习最优策略。确定马尔科夫决策过程需要定义决策过程的4个元素,即状态空间S、动作空间A、奖励函数R、转移概率P[18]。D3QN算法通过神经网络学习某个状态下的最佳动作策略,不需要再定义转移概率,因此在建立模型前需要确定三因素$\{S,A,R\}$。

建立状态空间合集$S=\{$OP_ID,TIME_ID,OBJ_VALUES$\}$。其中:OP_ID表示已加工订单对应的设备编号集合;TIME_ID记录所有设备已装配订单的加工时间;OBJ_VALUES表示每个设备上的已装配订单所获得的目标函数值集合。

动作空间A包含当前状态下的所有未加工订单索引。当动作执行时,设备从未加工订单集合中选择一个订单进行装配。执行动作后该订单的状态从未加工变成已加工状态,未加工订单集合也随之变化,动作动态变化采用掩码(mask)表示,当设备选择该订单后,未加工订单集合中将该订单的编号设置成-999,后面的动作执行过程则不会选择这个订单编号。

状态转移过程中环境会反馈奖励值,在本模型中设定为最小化目标函数值,但是在DRL中目标为寻求最大奖励,因此将执行动作的即时奖励

设置为上一个状态的目标函数 F_{t-1} 值减去当前状态值 F_t。

状态 s 执行动作 a 并转移到状态 s' 时获得的即时奖励为：

$$R_t = F_{t-1} - F_t \tag{9}$$

在一次迭代过程中执行完所有调度任务，总执行次数为 k 次，则训练过程中获得的累积奖励 R_{return} 为：

$$R_{\text{return}} = \sum_{t=1}^{k} R_t \tag{10}$$

2.3　算法求解流程

基于 D3QN 的训练原理，结合大规模定制的装配车间作业调度环境，得到的模型具体求解流程如图 1 所示。

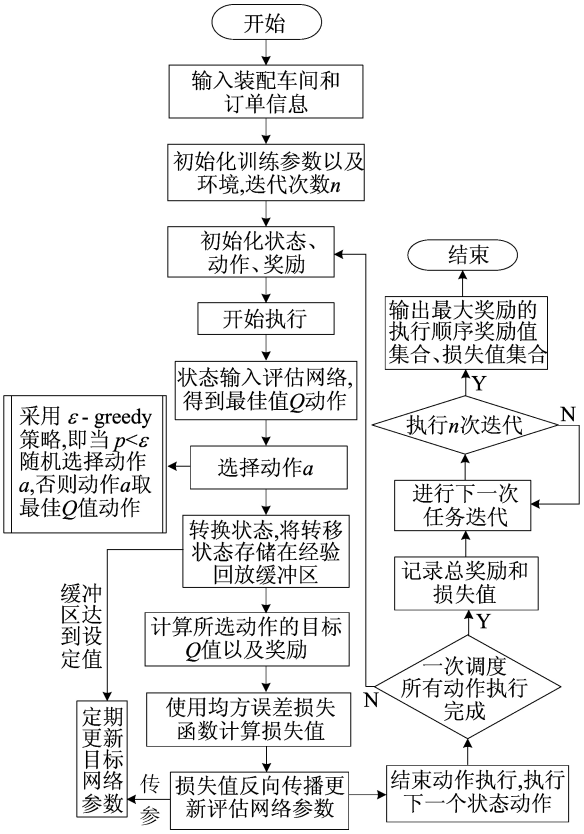

图 1　模型求解流程

3　算法仿真

3.1　实验

为了验证算法的有效性，搜集大规模定制装配车间的订单信息，以大规模定制为生产模式的某整车装配车间为例，进行实验验证。

该整车装配车间的汽车订单按照模块化进行划分，划分出的个性件见表 2 所列。

表 2 中，p_{ij} 为组件编号，$i=\{$外壳，轮毂，座椅，汽车音响$\}$，j 表示 4 个模块中不同类型的具体编号。

表 2　个性件具体分类

外壳	轮毂	座椅	汽车音响
白色 p_{11}	铝合金 p_{21}	真皮 p_{31}	高保真 p_{41}
黑色 p_{12}	碳纤维 p_{22}	仿皮 p_{32}	普通型 p_{42}
红色 p_{13}	钢质 p_{23}	织物 p_{33}	豪华型 p_{43}
银色 p_{14}			

在汽车订单中，除了包含汽车组件的具体要求外，还有订单交付期和订单数量等信息，具体见表 3 所列。

表 3　订单信息示例

订单编号	外壳颜色	轮毂材质	座椅材质	汽车音响	数量	交付日期
001	白色	铝合金	真皮	高保真	10	2023-04-10

根据订单信息随机生成 200 个订单规模的算例进行参数实验，并选取 3 个组装设备装配该订单。

基于实验调试要求以及文献[19]的实验结论，本实验设置 DRL 的学习率 l_r 为 10^{-6}，迭代次数为3 000，经验池 M_i 大小为 50 000，随机梯度下降采样样本大小为 128，神经网络更新频率为 200。

3.2　算法分析

经地 D3QW 算法训练后的总目标函数值如图 2 所示。

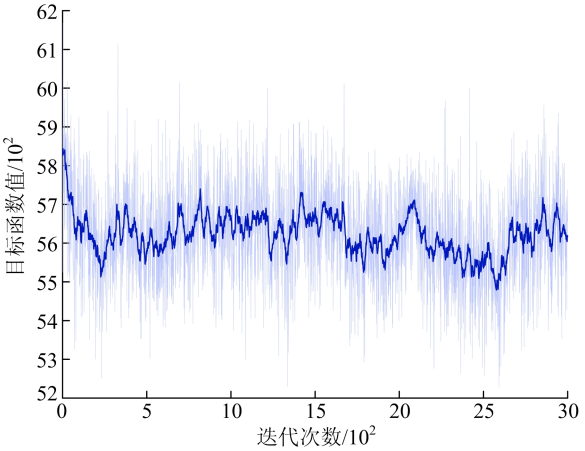

图 2　学习过程中目标函数值的变化

由图 2 可知：由于模型在前期训练过程中要储存训练数据，选择动作使用随机策略，而这些随

机策略可能不会有好的奖励,智能体需要不断地探索和尝试动作来找到环境中的所有可能性,从而获得更大的回报;当迭代 300 次左右时,由于动作选择策略开始从随机选取转化成选取 Q 值最大的动作,奖励开始有回报,并逐渐依靠好的策略实现目标函数值的减少。

对比现有装配车间所使用的按订单到达顺序进行组装的装配方式、使用 DQN 算法训练出的装配方式和使用 D3QN 算法训练出的装配方式的奖励值,结果见表 4 所列。

表 4 奖励值对比

指标	装配方式		
	按订单到达顺序	DQN 训练	D3QN 训练
交货期惩罚	9 850.537	8 908.100	8 714.090
组件相似度	55.600	55.900	57.500
总奖励值	−5 910.315	−5 344.853	−5 228.447

从表 4 可以看出,D3QN 算法训练出的装配顺序订单总交货期惩罚最小,订单组件总相似度最大,总奖励值最大。这意味着 D3QN 算法能够在降低提前或延期交货损失、最小化组件更换次数方面表现出较好的性能,从而降低库存取出成本和组件匹配成本。相较于现有的按照订单到达顺序进行组装的装配方式,使用 D3QN 算法可以获得更好的性能和效果。

比较 DQN 算法与 D3QN 算法的训练效果,如图 3 所示。

从图 3 可以看出,D3QN 的损失值比 DQN 的损失值更快地收敛。这说明 D3QN 通过引入目标网络的更新策略和使用不同的探索策略等,在函数逼近能力方面表现更优,这些改进使 D3QN 能够更准确地估计 Q 值,更好地平衡探索和利用,减轻过估计和不稳定性等问题,进而提高训练效果和收敛性。

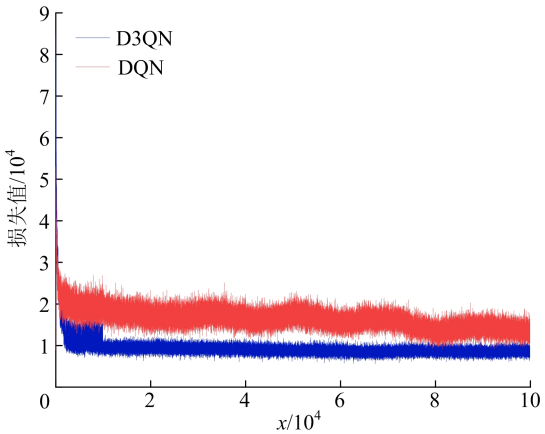

图 3 损失值对比

使用 D3QN 训练得到的最小目标函数值为 5 228.447,该值在迭代次数为 2 590 时出现,此次迭代的甘特图如图 4 所示。

从图 4 可以看出所有订单的具体安排顺序以及安排的设备序号。

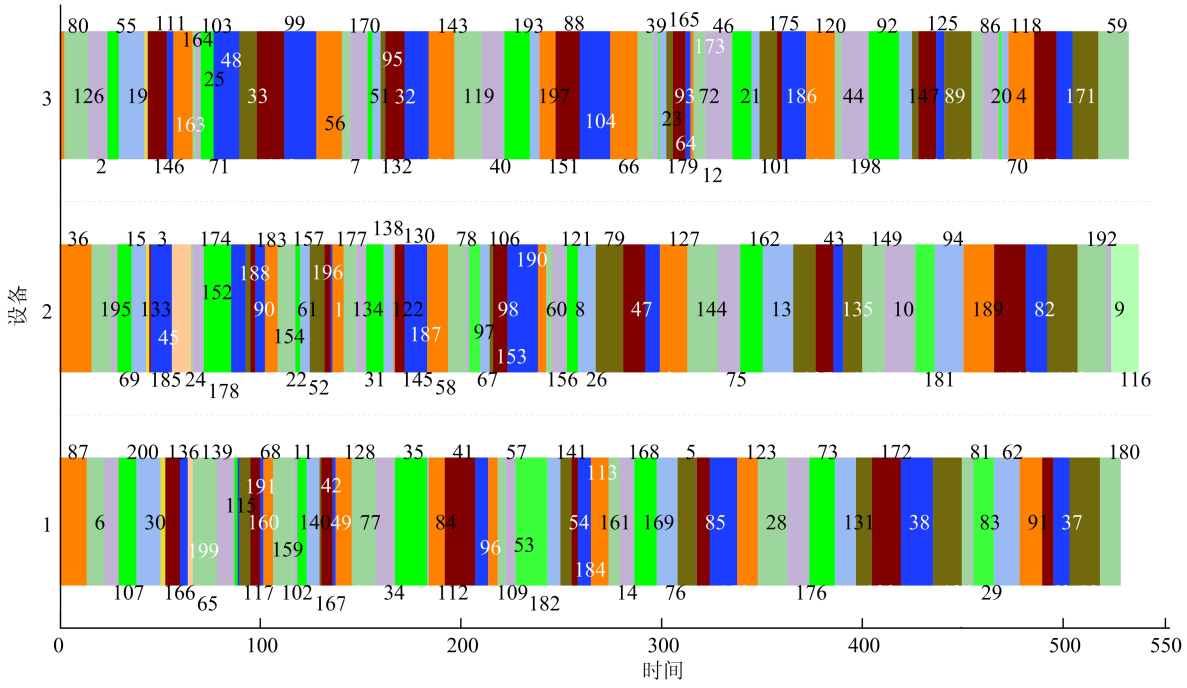

图 4 调度甘特图

4 结　论

本文基于大规模定制的订单装配调度特点，建立了优化模型，以最小的产品组件更换次数和最小的订单提前/拖期惩罚为目标，采用 D3QN 深度强化学习算法，并进行算法实验验证，最终得出结论如下：

1）设计了马尔可夫决策过程，该模型不仅能更好地描述调度环境，还能使智能体与环境进行交互。

2）订单的合理分配能够有效地满足交付期要求并且减少匹配成本和库存取出成本，提高生产效率。

3）所提出的 D3QN 算法可以获得较好的性能和效果，达到优化装配过程的目标，相较于现有的按订单到达顺序进行组装的装配方式，能够取得更好的结果；而且 D3QN 算法在训练过程中表现出更好的收敛性和效果，进一步证明了其改进的算法和网络结构的优势。

[参　考　文　献]

[1] 陈虹君，高伟. 大规模定制生产模式的研究现状及发展方向[J]. 现代商业，2016(26)：122-124.

[2] 李凡，何山，车晋伟，等. 大规模定制模式下动力锂电池生产计划策略与算法研究[J]. 制造业自动化，2021，43(1)：4-7，16.

[3] 周炳海，费芊然. 考虑能耗的混流设备物料配送多目标调度方法[J]. 东北大学学报(自然科学版)，2020，41(2)：258-264.

[4] ZHOU B，HE Z. A static semi-kitting strategy system of JIT material distribution scheduling for mixed-flow assembly lines[J]. Expert Systems with Applications，2021，184：115523.

[5] 娄高翔，蔡宗琰，刘清涛. 一种新型混合算法在混流装配调度中的应用[J]. 计算机工程与应用，2018，54(16)：254-259.

[6] 雷斌，刘同朝. 基于 PSO-GA 混合算法的转向架混流装配车间生产调度研究[J]. 现代制造工程，2020(7)：19-24.

[7] 李燚，唐倩，刘联超，等. 基于改进蚁群算法的汽车混流装配调度模型求解[J]. 中国机械工程，2021，32(9)：1126-1133.

[8] YAGMAHAN B. Mixed-model assembly line balancing using a multi-objective ant colony optimization approach[J]. Expert Systems with Applications，2011，38(10)：12453-12461.

[9] 解占新，闫玺铃，陆春月. 面向订单的混流设备多目标调度研究[J]. 机电工程，2021，38(5)：580-586.

[10] 李斌，陈立平，黄正东，等. 面向大规模定制的设备优化调度研究[J]. 中国机械工程，2005(24)：2198-2202.

[11] CAI C Z，KAN S L. Real-time scheduling of mixed model assembly line with large variety and low volume based on event-triggered simulated annealing (ETSA)[J]. Mathematical Problems in Engineering，2021，2021：6657506.

[12] 甘雅文，侯亮，徐昌华，等. 考虑产品切换的客车混流设备排序问题[J]. 计算机集成制造系统，2019，25(7)：1685-1694.

[13] ZHANG C，SONG W，CAO Z，et al. Learning to dispatch for job shop scheduling via deep reinforcement learning[J]. Advances in Neural Information Processing Systems，2020，33：1621-1632.

[14] 谭远良，吕佑龙，左丽玲，等. 基于强化学习的航天产品装配线投产排序研究[J]. 组合机床与自动化加工技术，2022(7)：160-164.

[15] 胡一凡，张利平，白雪，等. 深度强化学习求解柔性装配作业车间作业调度问题[J]. 华中科技大学学报(自然科学版)，2023，51(2)：153-160.

[16] 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.

[17] WANG H X，SARKEY B R，LI J，et al. Adaptive scheduling for assembly job shop with uncertain assembly times based on dual Q-learning[J]. International Journal of Production Research，2020，59(19)：5867-5883.

[18] 阮应君，侯泽群，钱凡悦，等. 基于深度强化学习的分布式能源系统运行优化[J]. 科学技术与工程，2022，22(17)：7021-7030.

[19] 乔东平，段绿旗，黎宏磊，等. 基于深度强化学习的作业车间作业调度问题优化[J]. 制造技术与机床，2023(4)：148-155.

（责任编辑　胡亚敏）