基于强化学习的智能车间调度策略研究综述*

王无双，骆淑云
（浙江理工大学 信息学院，杭州 310000）

摘　要：智能制造是我国制造业发展的必然趋势，而智能车间调度是制造业升级和深化“两化融合”的关键技术。主要研究强化学习算法在车间调度问题中的应用，为后续的研究奠定基础。其中车间调度主要包括静态调度和动态调度；强化学习算法主要包括基于值函数和 AC（Actor-Critic）网络。首先，从总体上阐述了强化学习方法在作业车间调度和流水车间调度这两大问题上的研究现状；其次，对车间调度问题的数学模型以及强化学习算法中最关键的马尔可夫模型建立规则进行分类讨论；最后，根据研究现状和当前工业数字化转型需求，对智能车间调度技术的未来研究方向进行了展望。

关键词：强化学习；动态调度；静态调度；作业车间调度；流水车间调度
中图分类号：TP181　　**文献标志码**：A　　**文章编号**：1001-3695(2022)06-002-1608-07
doi:10.19734/j.issn.1001-3695.2021.12.0637

Research on intelligent shop scheduling strategies based on reinforcement learning

Wang Wushuang, Luo Shuyun
(*School of Information Science & Technology, Zhejiang Sci-Tech University, Hangzhou* 310000, *China*)

Abstract: Intelligent manufacturing is an inevitable trend in the development of our country's manufacturing industry, and intelligent shop scheduling is a key technology for the integration of manufacturing upgrades and deepening. This paper mainly studied the application of reinforcement learning algorithms in shop scheduling problems, which layed the foundation for subsequent research. Shop scheduling mainly included static scheduling and dynamic scheduling; reinforcement learning algorithms mainly included value-based functions and Actor-Critic (AC) networks. First of all, this article described the research status of reinforcement learning methods on the two major issues of job shop scheduling and flow shop scheduling in general. Secondly, it classified the establishment rules of mathematical model of the shop scheduling problem and the most critical Markov model in reinforcement learning algorithms. Finally, according to the research status and the current needs of industrial digital transformation, it prospected the future research direction of intelligent workshop scheduling technology.

Key words: reinforcement learning; dynamic scheduling; static scheduling; job shop scheduling; flow shop scheduling

0　引言

车间调度问题是指如何在机器等资源有限的情况下，合理调度生产资源来安排车间生产任务，以满足一至多个优化目标的过程[1]。社会在不断发展，人民的消费水平也随之不断提高，各大企业在消费市场的竞争也愈演愈烈。有效的车间调度不仅可以提高生产量，还可以减少企业生产成本，提高客户满意度以及减少环境污染等。根据不同的生产特点，车间调度问题可以分为作业车间调度问题、流水车间调度问题以及开放车间调度问题。由于实际生产环境极为复杂，需要考虑各方面因素，所以出现了许多种基于以上调度问题的延伸问题，如柔性作业车间调度问题、混合流水车间调度问题等，它们往往更为复杂。车间调度问题还可以分为静态调度和动态调度两大类。静态调度即在开始生产前，已经获得了所有关于生产任务的信息且生产环境稳定。而动态调度是指在生产环境不确定的情况下进行调度，更符合实际生产情况[2]。在实际生产过程中，往往会发生诸如机器故障、紧急插单等突发状况。因此，动态车间调度比静态车间调度更为复杂，面向动态车间调度的研究也相对不成熟。

目前已有大量算法被应用于车间调度问题的研究中，较为常见的有数学规划方法、智能算法以及图与网络算法等[3]。其中，智能算法的应用最为成熟，如人工蜂群算法、蛙跳算法等[4,5]。近几年，强化学习应用于研究车间调度问题的优势引起了大量研究者的注意。强化学习以试错的方式进行学习，通过与环境交互获得奖励来指导动作，目标是获得最大的累积奖励。此外，强化学习还能够应对环境的不确定性，具有很强的适应性[6]。因此，无论是应用强化学习解决静态车间调度问题还是动态车间调度问题，都已经取得了令人欣喜的成果。由于作业车间以及流水车间的应用背景极为广泛，已经涵盖了大部分企业生产车间的特点，且目前应用强化学习解决开放车间调度问题的研究尚处于起步阶段。所以，本文系统地阐述了应用强化学习解决作业车间调度问题以及流水车间问题的研究现状，暂时未将开放车间调度问题考虑在内。

作业车间调度问题被定义为：一个加工系统有 M 台机器，要求加工 N 个工件，其中，每个工件完工都需要经过一定的工序加工。各工序的加工时间已确定，并且每个工件必须按照工序的先后顺序加工，工件所有工序只有唯一的加工机器。调度任务是安排所有作业的加工顺序，在满足约束条件的同时，使性能指标得到优化[7]。而如果工序加工所需要的资源是具备柔性的，即一道工序有多台机器可供选择，那么作业车间调度问题就拓展为柔性作业车间调度问题[8]。很显然，柔性作业车间调度问题更符合实际生产场景。

流水车间调度问题被定义为：N 个工件要在 M 台机器上加工，每个工件需要经过 M 道工序，N 个工件在 M 台机器上的加工顺序相同。工件在机器上的加工时间给定，要求确定每个

收稿日期：2021-12-08；**修回日期**：2022-02-07　**基金项目**：浙江理工大学基本科研业务费专项资金资助项目(2021Q026)
作者简介：王无双(1997-)，女，浙江绍兴人，硕士研究生，主要研究方向为强化学习、车间调度(wangwushuang_zstu@163.com)；骆淑云(1986-)，女，浙江金华人，讲师，硕导，博士，主要研究方向为工业互联网.

工件在每台机器上的最优加工顺序[9]。而如果工序加工所需要的资源具备柔性特征,那么流水车间调度问题就扩展为柔性流水车间调度问题,也称为混合流水车间调度问题[10]。如果规定 N 个工件的加工顺序对所有 M 台机器均相同,则称其为置换流水车间调度问题;如果允许在不同机器上改变工件的加工顺序,则称其为非置换流水车间调度问题。若每台机器加工任意两相邻工件时没有空闲时间,那么流水车间调度问题就扩展为无等待流水车间调度问题[11]。此外,若产品的生产主要包括加工阶段和装配阶段两个阶段,在每个阶段都要经过多台机器加工,该问题即为两阶段流水车间调度问题[12]。

根据迭代方式的不同可将强化学习算法分为基于值函数的强化学习算法、基于策略的强化学习算法以及基于 AC(Actor-Critic)的强化学习算法三种。由于基于策略的强化学习算法适用于求解具有连续动作空间的问题,而车间调度问题为组合优化问题,所以目前并没有研究者采用基于策略的强化学习算法解决车间调度问题。

本文将涉及在内的两种车间调度问题分为静态调度问题和动态调度问题,并对现有的研究成果按强化学习算法分类进行系统阐述并总结状态、动作以及奖励的设置方法。最后结合车间调度问题的发展趋势,分析并展望了该问题的未来研究方向。

1　作业车间调度

本章主要从静态调度和动态调度两个方面来阐述作业车间调度问题的研究现状,并根据强化学习算法的不同类别进行详细介绍。图 1 为本文所涉及到的作业车间问题分类框架图。

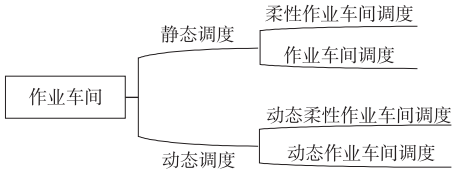

图 1　作业车间调度问题分类

Fig. 1　Classification of job shop scheduling problem

针对解决静态作业车间调度问题的研究成果,目前基于值函数的强化学习算法占主体研究导向,因此在该类方法下,本章首先介绍单独使用强化学习解决该问题的研究成果,包括单智能和多智能两类。其次介绍强化学习和其他算法相结合的混合算法如何解决该问题,包括其他算法赋能强化学习算法和强化学习算法赋能其他算法两种方式。表 1 为采用基于值函数的强化学习算法解决静态作业车间调度问题的研究成果汇总。最后,本章对采用基于 AC 的强化学习算法来解决静态作业车间调度问题的研究成果进行了分析说明。

针对动态作业车间调度问题的研究,动态环境的研究模型都有所差异,其中重点研究了工件随机到达、新作业插入等动态因素。本文主要按不同动态因素来分类阐述基于值函数的强化学习算法解决动态作业车间调度问题的研究成果,同时也对采用基于 AC 的强化学习算法来解决动态作业车间调度问题的成果进行了分析。表 2 为采用基于值函数的强化学习算法解决动态作业车间调度问题的研究成果汇总表。

表 1　基于值函数的强化学习算法解决静态作业车间调度问题

Tab. 1　Reinforcement learning algorithm based on value function to solve static job shop scheduling problem

参考文献	所用方案	解决的问题	优化目标	所用算法
[13]	单纯使用单智能体强化学习算法	柔性作业车间调度	最小化作业平均等待时间	Q 学习
[15]			最小化最大完工时间	时间差分
[14]		作业车间调度	最小化最大完工时间	Q 学习
[16]			最小化最大完工时间	DQN
[17]	单纯使用多智能体强化学习算法	柔性作业车间调度	最小化总延期时间和总完工时间	多智能体 DQN
[19]			最小化总延期时间和总完工时间	多智能体 DQN
[18]		作业车间调度	最小化总延期时间和总完工时间	多智能体 Q 学习
[20]	使用其他算法赋能强化学习算法	作业车间调度	最小化最大完工时间	DQN 结合 MEC 以及迁移学习
[21]			最小化最大完工时间	D3QN 结合析取图调度
[22]			最小化最大完工时间	结合 k 均值聚类和 Q 学习
[23]			最小化最大完工时间	强化学习结合析取图调度
[24]	使用强化学习算法赋能其他算法	柔性作业车间调度	最小化最大完工时间	Q 学习和遗传算法
[25]			最小化最大完工时间、机器总负荷及瓶颈机器负荷	强化学习结合非支配序列遗传算法

表 2　基于值函数的强化学习算法解决动态作业车间调度问题

Tab. 2　Reinforcement learning algorithm based on value function to solve dynamic job shop scheduling problem

参考文献	动态因素	解决的问题	优化目标	所用算法
[29]	机器故障	动态柔性作业车间调度	最小化延期时间	遗传算法与 Q 学习
[30]			最小化最大完工时间	多智能体 DQN
[31]	作业随机到达	动态作业车间调度	最小化最大完工时间	DQN
[32]			最小化延期时间	DQN
[33]			最小化过早完工和延期的惩罚	多智能体 Q 学习
[34]	新作业插入	动态柔性作业车间调度	最小化延期时间	DDQN
[35]		动态作业车间调度	最小化延期时间和最大化机器平均利用率	DDQN
[36]	机器故障和作业随机到达	动态作业车间调度	最小化工件平均流动时间	可变邻域搜索与 Q 学习
[37]	作业随机到达和产品随机选择	动态作业车间调度	最小化过早完工时间	Q 学习

1.1　静态调度

1.1.1　基于值函数的强化学习算法

目前,有部分研究团队单纯采用基于值函数的强化学习算法来解决静态作业车间调度问题。Bouazza 等人[13]首先将部分柔性作业车间调度问题分解成机器分配子问题和机器上作业分配子问题,其次使用 Q 学习算法,其中智能体的动作为选择机器分配规则和选择机器上工件的调度规则。针对该问题,王维祺等人[14]在利用 Q 学习算法时将策略选择的优先级作为学习对象,并分别设定五个基本动作对应五种不同的状态,策略为动作和状态的组合。为了更符合实际应用场景,Martins 等人[15]考虑了双资源约束的柔性作业车间调度问题,即同时考虑分配机器和工人资源,并最小化最大完工时间。为增强算法的鲁棒性,减少机器和工件数量规模对算法性能的影响,Samsonov 等人[16]提出了一种新的动作空间设计方法,其能够使动作空间的规模不受作业和作业工序数量的影响,同时不受连续空间和离散空间的限制。该动作空间设置方法能够将解决问题的方法从基于离散动作空间的 DQN(deep Q-network)算法进一步拓展到策略梯度(policy gradient,PG)、深度确定性

策略梯度(deep deterministic policy gradient,DDPG)等基于连续动作空间的强化学习算法。

强化学习除了单智能体学习,还有多智能体学习,虽然其有诸如性能不稳定、维度爆炸等缺陷,但更适用于实际工厂生产调度场景。文献[17]提出了一种多智能体深度强化学习算法解决具有并行不相关机器的作业车间调度问题,以最小化拖期时间和最大完工时间为优化目标。实验结果表明,其在大规模问题上的性能要比混合整数线性规划算法更有优势。Méndez-hernández 等人[18]也采用类似方法解决同样的问题。在该两阶段优化方法中,智能体在第一阶段作为独立单元优化各自代表的目标,在第二阶段协作寻找折中的解决方案。Lang 等人[19]将柔性作业车间调度问题分为两个子问题,并训练了两个 DQN 智能体,其中一个负责工序顺序的选择,另一个负责将工件分配给机器。

由于作业车间调度问题具有高度复杂性,且强化学习自身存在收敛速度慢、易陷入局部最优等缺陷,很多研究者尝试结合其他算法来解决以上问题。为使调度具有实时性,Moon 等人[20]提出了一种基于协同边缘计算框架的智能制造工厂生态系统架构,并使用 DQN 算法来解决其作业车间调度问题。该框架还引用了迁移学习,通过将历史任务上所学知识迁移到新任务中来加快算法在新环境中的学习速度。实验结果表明,该方法在不同参数规模下都比传统方法具有更好的收敛效果。Han 等人[21]提出了一种基于析取图调度的深度强化学习框架,并进一步提出了带优先经验重放的双深度 Q 学习(dueling double DQN,D3QN)算法,其主要基于竞争网络结构,并通过将状态表示为多通道图像的方式来更好地提取状态特征。实验证明,针对小规模问题,该算法能够获得最优解;对于大规模问题,其性能优于任何单一的启发式规则,且与遗传算法相当。该研究团队又提出了一种结合编码器网络的调度方法[22],并采用策略梯度算法来优化其参数。实验表明,训练模型的效果比启发式算法更好。Lara-Cárdenas 等人[23]结合 K 均值聚类和 Q 学习算法来解决作业车间调度问题。实验结果表明,该算法的性能优于一些基于最短加工时间规则和最大作业剩余时间规则的启发式算法。

由于强化学习能够通过与环境交互,在试错中学习正确行为,研究学者将其应用于解决智能优化算法中的参数调节和种群多样性控制等难题。Chen 等人[24]结合 Q 学习算法和 Sarsa 算法去改进遗传算法来解决柔性作业车间调度问题,其根据种群的当前状态(种群的平均适应度值、种群多样性和最好个体的适应度值),利用 Q 学习和 Sarsa 算法自适应调整交叉和变异的概率。针对柔性作业车间调度问题,非支配排序遗传算法存在易陷入局部的硬伤,尹爱军等人[25]融合多个多样性指标,利用强化学习动态优化种群迭代过程中的拆分比例参数以保持种群多样性。

1.1.2 基于 Actor-Critic 的强化学习算法

目前,应用基于 AC 的强化学习算法来解决静态作业车间调度问题的研究尚处于起步阶段,仅有三个研究团队分别采用 AC、近端策略优化(proximal policy optimization,PPO)以及多智能体 PPO 算法来解决该问题。Liu 等人[26]提出了包括卷积层和全连接层的 AC 网络去解决该问题。为加快收敛速度,还提出了一种结合异步更新和 DDPG 算法的并行训练算法。实验结果表明,针对静态调度环境,该方法比传统启发式算法性能更好,同时也能推广到动态调度环境中应用。Park 等人[27]结合图神经网络和 PPO 算法来解决该问题,采用图神经网络来学习嵌入表示为图的作业车间调度问题空间结构的节点特征,并生成将嵌入节点特征映射到最佳调度动作的调度策略。实验表明,该框架在训练得到模型后,无须进一步训练即可应用于新的作业车间调度问题中,大大节省了重新训练所需的时间。由于作业延误和过早完工都会带来不可忽略的生产成本,Roesch 等人[28]在综合考虑生产成本与能源成本的前提下,利用多智能体 PPO 算法来解决该问题,其中每个智能体都代表一台机器且必须处理一定数量的作业。

1.2 动态调度

1.2.1 基于值函数的强化学习算法

针对机器故障的动态作业车间调度问题,Zhao 等人[29]提出了一种改进的 Q 学习算法。当机器发生故障时,Q 学习智能体能够同时选择该处理的工序以及可替代的机器。机器发生故障前的初始调度方案由遗传算法获得。实验结果表明,与单一调度规则相比,所提方案能够减少频繁动态环境中的作业延迟时间。Bär 等人[30]使用了多智能体 DQN 算法来解决动态柔性作业车间调度问题,将每个产品设置为智能体并共用一个经验回放池。这些智能体能够在训练中学会合作,考虑其他智能体的需求以实现优化目标。

针对作业随机到达的动态作业车间调度问题,Luo 等人[31]提出了一个具有探索循环和利用循环的双循环 DQN 算法。该算法集成了探索循环的全局探索能力以及利用循环的局部收敛能力,可促进 DQN 算法找到问题的全局最优解。Turgut 等人[32]也采用 DQN 算法来解决该问题,以最小化作业延误时间。实验结果表明,该方案有两个启发式调度规则,即最短处理时间和最早到期日更有效。在实际生产中,过早完工会给企业带来库存压力,延误工期会影响客户满意度。考虑到以上两点,Wang[33]建立了一个基于多智能体的动态调度系统模型,将机器、缓冲区、状态和作业设为智能体,并使用加权 Q 学习算法来确定作业在机器上的加工顺序。生产车间的动态环境会引起系统状态变化,从而导致状态空间巨大。针对该问题,作者定义了四个状态特征,并通过聚类的方法降低了状态空间维度。此外,为避免传统策略中的盲目搜索,还提出了一种动态贪婪搜索策略。

针对新作业插入的动态作业车间调度问题,为减少作业延误带来的成本,Luo[34]应用了 DQN 算法,提取了七个取值于[0,1]的通用特征来表示每个重调度点的状态,并将动作设计为六个可选复合规则以确定下一个要处理的工序和分配给它的机器。针对该问题,Luo 等人[35]还提出了双层 DQN 在线重调度框架,该框架包含了两个基于 DQN 的智能体。上层 DQN 用来控制下层 DQN 的临时优化目标,在每个重调度点,它将当前状态特征作为输入,并根据优化目标来指导下层 DQN 的行为。下层 DQN 将状态特征和从上层 DQN 传递的优化目标作为输入,将每个调度规则的 Q 值作为输出。基于该 Q 值,可选择每个重调度点上最可行的调度规则。

针对机器故障和作业随机到达的动态作业车间调度问题,Shahrabi 等人[36]使用可变邻域搜索算法来解决该问题,以最小化完工时间。针对可变邻域搜索算法易陷入局部最优等缺陷,使用 Q 学习算法在每个重调度点上更新它的参数。实验结果表明,该方案比传统启发式调度规则更有效。

针对作业随机到达和产品随机选择的动态作业车间调度问题,为减少提前完工带来的库存压力,Kardos 等人[37]应用了多智能体 Q 学习算法来解决该问题,将每个产品设为智能体并能够根据实时信息在每个生产步骤选择机器。与标准调度规则的比较表明,该方案具有更好的性能。

1.2.2 基于 Actor-Critic 的强化学习算法

目前采用基于 AC 的强化学习算法来解决动态作业车间调度问题还处于初步探索阶段。Wang 等人[38]考虑了诸如机器故障、工件返工等各种动态因素,尝试利用 PPO 算法解决该问题,以最小化最大完工时间。不同于一般的研究方法,作者将状态定义为三个矩阵,分别为作业处理状态矩阵、机器指定矩阵和工序的处理时间矩阵。经实验证明,所提方案的性能优

于传统启发式规则以及遗传算法,且在一定程度上可以实现自适应调度。

2　流水车间调度

本章主要从静态调度和动态调度两个方面来阐述流水车间调度问题的研究现状,并根据强化学习算法的不同类别进行详细介绍。图 2 为本文所涉及到的流水车间调度问题分类框架图。

针对解决静态流水车间调度问题的研究成果,目前基于值函数的强化学习算法占主体研究导向。在该类方法下,本章首先介绍单独使用强化学习解决该问题的研究成果,将问题根据是否考虑自动导引运输车(automated guided vehicle,AGV)分为两类。其次介绍如何利用强化学习和其他算法相结合的混合算法来解决该问题。表3为采用基于值函数的强化学习算法解决静态流水车间调度问题的研究成果汇总表。最后,本章对采用基于 AC 的强化学习算法来解决静态流水车间调度问题的研究成果进行了分析说明。

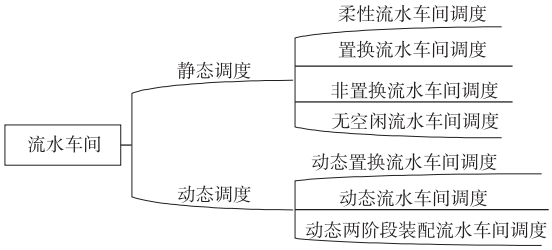

图 2　流水车间调度问题分类

Fig. 2　Classification of flow shop scheduling problem

针对动态流水车间调度问题,本章对采用基于值函数和基于 AC 的强化学习算法来解决该问题的研究成果进行了分析。

表 3　基于值函数的强化学习算法解决静态流水车间调度问题

Tab. 3　Reinforcement learning algorithm based on value function to solve static flow shop scheduling problem

参考文献	所用方案	解决的问题	优化目标	所用算法
[39]	单纯使用强化学习算法	柔性流水车间调度	最小化各工件加工时间	Q 学习
[40]		置换流水车间调度	最小化最大完工时间	自适应 Q 学习
[41]			最小化最大完工时间	Q 学习
[42]		非置换流水车间调度	最小化机器空闲时间	TD
[43]		考虑 AGV 的流水车间调度	最小化延期时间以及最大完工时间	Q 学习
[44]			最小化最大完工时间	多智能体 Q 学习
[45]	强化学习算法结合其他算法	柔性流水车间调度	最小化各作业平均延期时间	ATCS 规则和 Q 学习
[46]		置换流水车间调度	最小化最大完工时间	自适应 Q 学习、局部搜索算法
[47]			最小化最大完工时间	强化学习和迭代贪婪
[48]		无空闲流水车间调度	最小化最大完工时间	可变邻域搜索和 Q 学习

2.1　静态调度

2.1.1　基于值函数的强化学习算法

目前,有部分研究团队单纯采用基于值函数的强化学习算法来解决静态流水车间调度问题。Han 等人[39]首次提出使用 Q 学习算法来解决混合流水车间问题,在算法中采用玻尔兹曼探索策略来平衡探索和利用,并以汽车发动机金属加工厂的实例来对该算法进行验证。在确保复杂度相同的情况下,该算法的性能优于遗传算法,且该算法的收敛速度比人工免疫算法更快。Reyna 等人[40]采用了自适应 Q 学习算法来解决具有序列相关生产时间、机器初始化准备时间的置换流水车间调度问题,将状态定义为作业优先级关系,将动作定义为更改作业优先级。张东阳等人[41]同样采用 Q 学习算法来解决该问题。不同的是,该算法将状态定义为作业序列,将动作定义为选择可选的工件,最后用 OR-Library 提供的标准算例进行仿真实验,结果表明,相较于其他智能算法以及启发式规则,Q 学习算法的寻优能力更好。肖鹏飞等人[42]提出了深度时间差分网络算法来解决非置换流水车间调度问题。该算法采用深度神经网络来拟合值函数,用 TD 算法来训练网络模型,将启发式调度规则设为动作并结合 AC 网络结构为每次调度决策选取最优的组合行为策略。实验证明,相较于群智能算法,该算法的性能更优。

在实际生产中,时常会用到智能运输小车等工具进行物料运输。针对该现象,部分研究者将柔性搬运系统结合到流水车间调度问题中。Xue 等人[43]应用 Q 学习算法来解决流水车间中 AGV 的调度问题,将总完工时间最小化作为优化目标。考虑到完工时间主要受 AGV 等待时间与作业等待时间的影响,将 AGV 作为智能体,根据系统当前情况来决定所需要完成的任务,并设计改进的 ε-greedy 方法来平衡探索和利用。实验结果表明,在问题规模较大的情况下,该算法的性能优于多智能体算法。Arviv 等人[44]研究了具有两个机器人的流水车间调度问题,定义了四种机器人协作方式,并使用双重 Q 学习算法,将两个机器人设为智能体,并给它们分配了不同奖励。其中一个机器人负责最小化机器空闲时间,另一个则负责最小化作业等待时间,机器人之间可通过交换奖励值来共享信息。为验证算法性能,用快速、中速以及慢速机器人进行仿真实验,结果表明,两个快速机器人之间的完全协作能够取得最佳效果。

流水车间调度问题是一个 NP 难问题,解空间十分庞大,且强化学习存在易陷入局部最优等缺陷,许多研究者尝试结合其他算法来解决该问题。考虑准备时间的直观延误成本(apparent tardiness cost with setups,ATCS)规则的参数能够极大地影响该规则的性能,针对该问题,Heger 等人[45]提出采用 Q 学习算法来自主调整 ATCS 规则中的 k_1 和 k_2 值。与其他研究成果所能得到的最佳 k 值相比较,其得到的 k 值能将平均延迟降低 5%。针对置换流水车间调度问题,Yunior 等人[46]采用启发式算法来生成作业的初始排序,并结合自适应 Q 学习算法以及局部搜索算法来解决该问题。在实验中,将该算法与包括粒子群算法在内的八种其他算法进行比较,结果表明,该方案能在较短时间内得到更高质量的解。针对传统模型受问题规模影响而难以扩展的缺陷,王凌等人[47]设计了一种新的编码网络来对问题进行建模,通过深度强化学习算法训练模型来获得该问题的初始调度解,并采用带反馈机制的迭代贪婪算法来继续优化该初始调度解以获得最终调度解。Öztop 等人[48]采用可变邻域搜索算法来解决无空闲流水车间调度问题,并采用 Q 学习算法来自适应地调节可变邻域搜索算法的参数。实验结果表明,该算法的性能优于传统迭代贪婪算法。

2.1.2　基于 Actor-Critic 的强化学习算法

目前,在基于 AC 的强化学习算法中,仅有两个研究团队分别使用 PPO 以及 AC 算法来解决静态流水车间问题。Zhu 等人[49]首次采用 PPO 算法来解决具有相同并行机的混合流水车间调度问题,将最小化最大完工时间作为优化目标,并在真实实例和不同规模的随机实例上测试了该算法的性能。实验结果表明,在晶片酸洗实例上,该算法的性能优于遗传算法。在随机生成的实例上,该算法的性能优于其他启发式调度规

则。针对置换流水车间调度问题,Pan 等人[50]提出了一种异构网络深度强化学习模型,其中包括长短期内存网络(long short term memory,LSTM)和注意力网络。实验结果表明,该模型在较小规模问题上的性能优于传统启发式算法和其他相同结构的深度强化学习模型。

2.2 动态调度

2.2.1 基于值函数的强化学习算法

目前采用基于值函数的强化学习算法解决动态流水车间调度问题的研究还处于起步阶段。Yang 等人[51]为了实现动态调度的实时性以及智能决策,首次提出了利用深度强化学习来求解考虑新作业到达的动态置换流水车间调度问题,采用 A2C(advantage actor-critic)算法来训练网络模型。实验结果表明,该方案在解决方案质量、CPU 计算时间以及泛化能力等方面都明显优于传统元启发式算法。此外,Yang 等人[52]还采用 DDQN 来解决该问题,并通过大量的实例来进行训练,其效果优于一些经典的调度规则。Wang 等人[9]应用多智能体 Q 学习算法来解决该问题,将每个机器设为智能体,并在实验中验证了该算法的性能,却发现所提方案在实际应用中的缺陷。此后作者分析了强化学习的优缺点,针对原有算法的缺陷设计了两种改进策略。

2.2.2 基于 Actor-Critic 的强化学习算法

目前仅有一个研究团队采用基于 AC 的强化学习算法来解决动态两阶段装配流水车间调度问题。两阶段装配流水车间调度问题广泛存在于消防车制造、空调装配和船舶生产等制造业。两阶段流水车间的生产主要包括加工阶段和装配阶段两个阶段。产品先在加工阶段的多台专用机器上加工,而后被装配阶段的多台装配机器组装成成品。Lin 等人[12]采用 PPO 算法来解决动态两阶段装配流水车间调度问题,将最小化总延迟作为优化目标。在实验中,将单一的调度规则与所提方案进行比较。实验结果表明,无论生产订单的规模大小,PPO 算法所得调度方案的平均总延迟时间总是低于其他调度规则。

3 车间调度问题数学模型分析

车间调度问题具有十分广泛的工业应用背景,因此,需要了解其实际背景,明确其实际意义,以数学思想来包容问题的精髓。本章首先给出流水车间以及作业车间调度问题的一般假设和约束,再根据静态和动态生产环境分类来分析数学模型。

流水车间以及作业车间的特征相似,因此,这两个车间调度问题的假设以及约束条件基本一致。假设主要有以下几点:a)不允许作业抢占,一旦作业开始在一个机器上处理,该处理过程必须不间断直至完成;b)作业在每台机器上的处理时间已知;c)所有资源在零时刻可用;d)针对非柔性车间,一个作业的某道工序只能由一台机器加工。针对柔性车间,一个作业的某道工序可由多种类型的机器加工。约束条件主要有以下几点:a)同一作业的工序有先后顺序,如第一道工序必须在第二道工序开始之前完成;b)一台机器在同一时刻只能加工一个作业;c)一个作业在同一时刻只能在一台机器上处理。

针对静态生产环境下车间调度问题的数学模型,有两种比较传统的建模方式。一种是根据约束和假设建立混合整数线性规划模型,该模型的目标函数是线性的,约束条件也是线性的,而有部分或所有决策变量必须是整数。另一种是依据图论建模。该种建模方式一般都是采用有向图来建模,具体方式如下:$G=(V,C\cup D)$,其中 V 表示对应作业所有工序的一组顶点,C 是连接节点之间边的集合,表示同一作业的两个连续工序之间的优先级约束,D 为另一组连接节点的边,表示同一台机器上的任务顺序。动态生产环境下的数学模型较静态环境有一些改动,需给定实时事件发生的时间以及在此期间改动的模型参数。这些参数包括作业数量、可运作的机器、工序的处理时间以及供需之间的优先级关系等。

基于模型中的多目标函数,有以下两种较为常用的处理方式:a)直接以加权和的形式将多目标转换为单目标;b)根据帕累托法则进行求解。具体为每个目标分配一个智能体,采用多智能体算法来分阶段解决问题。在第一阶段,智能体作为独立单元,每个单元优化各自的目标;在第二阶段,智能体之间相互合作为所有目标组合找到最佳解决方案。

4 马尔可夫决策过程模型分析

虽然早在二十多年前就已经有研究者应用强化学习算法对车间调度问题进行了研究,但就目前的研究状况而言,该研究还处于不成熟阶段。强化学习算法虽然优势突出,但也伴随着收敛困难、难以平衡探索与利用等问题[53]。马尔可夫决策过程(Markov decision process,MDP)的模型建立好坏直接决定了强化学习算法的性能。因此,本章主要分析利用强化学习算法解决车间调度问题时如何建立 MDP 模型,侧重于阐述状态、动作和奖励三个要素的设置规律。图 3 为采用强化学习算法解决车间调度问题时,MDP 模型三要素的设规律汇总图。

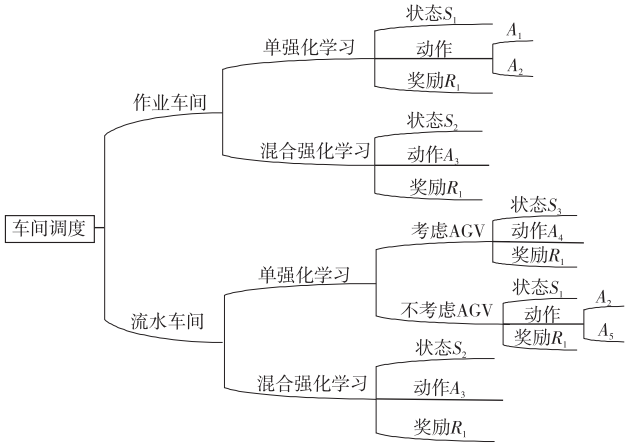

图 3 MDP 三要素分类设置规则

Fig. 3 MDP three-element classification setting rules

针对作业车间调度问题,单独的强化学习算法和优化群智能算法的强化学习算法在状态、动作和奖励设置上有所区别。一方面,对于单独的强化学习算法,状态一般定义为 S_1:工件和机器的特征,如工件在机器上所需的处理时间、工件的完成进度(已完成的工序数)和机器利用率等。动作一般为:a)A_1:选择下一步要处理的工序以及对应的机器;b)A_2:选择调度规则,通常指先进先出(first in first out,FIFO)等传统启发式调度规则。奖励一般为 R_1:与具体的优化目标有关。另一方面,对于优化群智能算法的强化学习算法,状态一般设置为 S_2:种群特征,如多样性、最大适应度值、平均适应度值等。动作一般为 A_3:要优化的参数值,奖励同样也依赖于优化目标。但也有研究学者没有遵循以上规则设计 MDP。Samsonov 等人[16]设计了一种新的动作空间,其能够保证动作空间的规模不受作业和作业工序数量的影响,并将动作设置为选择一段相对持续时间,随后利用最大最小处理时间将被选择的相对持续时间映射回绝对持续时间,最后选择与该绝对持续时间具有最接近处理时间的工序。

针对流水车间调度问题,单独的强化学习算法和优化群智能算法的强化学习算法在状态、动作和奖励设置上两者有所区别。一方面,对于单独的强化学习算法,MDP 的建模主要取决于是否引入 AGV。针对考虑 AGV 的流水车间调度问题,状态一般设置为 S_3:目前所有 AGV 所处的情况。动作一般为 A_4:选择机器的位置,因为 AGV 需要将上一台机器处理完毕的工

件运送到下一台机器,所以需要指示 AGV 移动轨迹。奖励同样也依赖于优化目标。针对不考虑 AGV 的流水车间调度问题,状态与奖励的设置规则与上述单独使用强化学习解决作业车间调度问题一致,但其动作的设置规则有所不同:a)A_2:选择调度规则;b)A_5:选择一个工件。另一方面,对于优化群智能算法的强化学习算法,MDP 的设置规则与上述使用强化学习优化其他智能算法来解决作业车间调度问题一致。但也有研究学者没有遵循以上规则设计 MDP。针对置换流水车间调度问题,王凌等人[47]将动作设置为直接给出各工件的加工顺序。针对混合流水车间问题,Zhu 等人[49]将状态设置为工件的加工顺序,通过初始化状态先设置一个工件的加工顺序,并将动作设置为选择两个工件并交换其位置。

5　结束语

智能车间调度是依据群智能优化算法、强化学习算法等来对车间产品的生产流程进行合理的调度和规划。随着企业数字化转型的需求与日俱增,智能生产和大型个性化定制等智能服务需求要求企业具有智能化的生产调度流水线,能够应对各种突发状况并能够高效完成订单任务。为了在消费市场中保持竞争力,车间调度是工厂在运营层面面临的最重要问题之一。下面从五个方面来阐述可行的车间调度的未来研究方向:

a)解决方案落地化。目前利用强化学习方法来解决车间调度问题的研究还处于不成熟阶段,大多数还停留在理论研究层面。由于生产调度是企业生产运营的关键环节,要提高生产效率还需要将算法运用到实际生产系统中。信息技术和运营技术融合是从理论转换成系统,从知识转换成生产力的必然趋势。因此,生产调度问题还需要从实际系统部署层面进行更深一步的研究。

b)多目标车间调度。目前的研究成果在解决车间调度问题时,所要优化的目标基本都是最大完工时间最小化,过于单一。企业的实际需求还需要考虑到节能减排、订单延误、库存压力等问题。因此,未来在对车间调度问题的研究中,可以考虑更倾向于多目标优化。

c)采用混合算法。单一智能算法解决车间调度问题已经不能满足其多样性需求,因此混合算法成为解决该问题的新宠,其可以弥补单一算法存在的局限性。目前已有一些学者结合强化学习算法和遗传算法来解决车间调度问题,并取得了一定的成效[24]。未来可以考虑结合其他领域算法,如机器学习算法、博弈论等。

d)考虑工人因素。目前的研究仅局限于优化机器资源,事实上机器需要由工人来操作,工人资源的分配也会对生产效率产生极大影响。因此,未来需要考虑机器与工人双资源约束的车间调度问题。

e)基于强化学习算法的改进。强化学习算法以试错的方式进行学习,如果无法平衡其探索和利用,就极易陷入局部最优解。在未来,应设计出更好的探索和利用方案。此外,大多数研究者倾向于应用 Q 学习,DQN 等传统强化学习算法来解决车间调度问题。当前,已设计出了很多其他性能更优的算法。因此,研究者可以尝试将新算法应用于车间调度问题。

参考文献:

[1] 罗哲,夏余平,米双山. 典型车间调度问题的分析与研 究[J]. 科技创新与应用,2020(9):60-61. (Luo Zhe, Xia Yuping, Mi Shuangshan. Analysis and research on typical workshop scheduling problems [J]. Technology Innovation Application,2020(9):60-61.)

[2] Shvalika C, Silva T, Karunanandaa A. Reinforcement learning in dynamic task scheduling: a review[J]. SN Computer Science, 2020, 1(6):1-17.

[3] Cebi C, Atac E, Sahingoz O K. Job shop scheduling problem and solution algorithms: a review[C]//Proc of the 11th International Conference on Computing, Communicationand Networking Technologies. Piscataway, NJ: IEEE Press, 2020:1-7.

[4] 吴锐,郭顺生,李益兵,等. 改进人工蜂群算法求解分布式 柔性作业车间调度问题[J]. 控制与决策,2019,34(12):2527-2536. (Wu Rui, Guo Shunsheng, Li Yibing, *et al*. Improved artificial bee colony algorithm for distributed and flexible job-shop scheduling problem[J]. Control and Decision, 2019, 34(12):2527-2536.)

[5] 雷德明,杨冬婧. 基于新型蛙跳算法的低碳混合流水车间 调度[J]. 控制与决策,2020,35(6):1329-1337. (Lei Deming, Yang Dongjing. A novel shuffled frog-leaping algorithm for low carbon hybrid flow shop scheduling[J]. Control and Decision, 2020, 35(6): 1329-1337.)

[6] 马骋乾,谢伟,孙伟杰. 强化学习研究综述[J]. 指挥控制与仿真,2018,40(6):68-72. (Ma Chengqian, Xie Wei, Sun Weijie. Summary of reinforcement learning research [J]. Command Control and Simulation, 2018, 40(6):68-72.)

[7] Zhang Jian, Ding Guofu, Zou Yisheng, *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.

[8] Xie Jin, Gao Liang, Peng Kunkun, *et al*. Review on flexible job shop scheduling[J]. IET Collaborative Intelligent Manufacturing, 2019, 1(3):67-77.

[9] Wang Jinzhi, Qu Shuhui, Wang Jie, *et al*. Real-time decision support with reinforcement learning for dynamic flowshop scheduling[C]// Proc of European Conference on Smart Objects, Systems and Technologies. [S. l.]: VDE Press, 2017:1-9.

[10] 李颖俐,李新宇,高亮. 混合流水车间调度问题研究综 述[J]. 中国机械工程,2020,31(23):2798-2813,2828. (Li Yingli, Li Xinyu, Gao Liang. Summary of research on scheduling problems of mixed flow shop [J]. China Mechanical Engineering, 2020, 31(23):2798-2813,2828.)

[11] Singh H, Oberoi J S, Singh D. Multi-objective permutation and non-permutation flow shop scheduling problems with no-wait: a systematic literature review[J]. RAIRO-Operations Research, 2021, 55(1):27-50.

[12] Lin Xin, Chen Jian. Deep reinforcement learning for dynamic scheduling of two-stage assembly flowshop[C]//Proc of International Conference on Swarm Intelligence. Berlin: Springer, 2021:263-271.

[13] Bouazza W, Sallez Y, Beldjilali B. A distributed approach solving partially flexible job-shop scheduling problem with a Q-learning effect [J]. IFAC-PapersOnLine, 2017, 50(1):15890-15895.

[14] 王维祺,叶春明,谭晓军. 基于 Q 学习算法的作业车间 动态调度[J]. 计算机系统应用,2020,29(11):218-226. (Wang Weiqi, Ye Chunming, Tan Xiaojun. Dynamic job shop scheduling based on Q-learning algorithm[J]. Computer Systems & Applications, 2020, 29(11):218-226.)

[15] Martins M S E, Viegas J L, Coito T, *et al*. Reinforcement learning for dual-resource constrained scheduling [J]. IFAC-PapersOnLine, 2020, 53(2):10810-10815.

[16] Samsonov V, Kemmerling M, Paegert M, *et al*. Manufacturing control in job shop environments with reinforcement learning[C]//Proc of the 13th International Conference on Agents and Artificial Intelligence. [S. l.]: Portugal: Scitepress-Science and Technology Publications Press, 2021:589-597.

[17] Wemelsfelder M. Approximating optimal solutions for job shop scheduling problems with unrelated machines in parallel using generalizable deep multi-agent reinforcement learning[D]. Amsterdam: University of Amsterdam, 2020.

[18] Méndez-hernández B M, rodríguez-bazan E D, Martinez-Jimenei Y, *et al*. A multi-objective reinforcement learning algorithm for jssp[C]// Proc of International Conference on Artificial Neural Networks. Berlin: Springer, 2019:567-584.

[19] Lang S, Behrendt F, Lanzerath N, *et al*. Integration of deep reinforcement learning and discrete-event simulation for real-time scheduling of a flexible job shop production[C]//Proc of Winter Simulation Conference. Piscataway, NJ: IEEE Press, 2020:3057-3068.

[20] Moon J, Yang M, Jeong J. A novel approach to the job shop scheduling

problem based on the deep Q-network in a cooperative multi-access edge computing ecosystem[J]. Sensors,2021,21(13):article ID 4553.
[21] Han Baoan, Yang Jianjun. Research on adaptive job shop scheduling problems based on dueling double DQN[J]. IEEE Access,2020,8:186474-186495.
[22] Han Baoan, Yang Jianjun. A deep reinforcement learning based solution for flexible job shop scheduling problem[J]. International Journal of Simulation Modelling (IJSIMM),2021,20(2):375-386.
[23] Lara-cárdenas E, Silva-gálveza, Optiz-Bayliss J C, *et al*. Exploring reward-based hyper-heuristics for the job-shop scheduling problem [C]//Proc of IEEE Symposium Series on Computational Intelligence. Piscataway, NJ:IEEE Press,2020:3133-3140.
[24] Chen Ronghua, Yang Bo, Li Shi, *et al*. A self-learning genetic algorithm based on reinforcement learning for flexible job-shop scheduling problem[J]. Computers & Industrial Engineering,2020,149:article ID 106778.
[25] 尹爱军,闫文涛,张厚望. 面向多目标柔性作业车间调度 的强化学习 NSGA-II 算法[J/OL]. 重庆大学学报,2021. (2021-05-12)[2022-01-10]. http://kns. cnki. net/kcms/detail/50. 1044. N. 20210511. 1913. 0 10. html. (Yin Aijun, Yan Wentao, Zhang Houwang. Reinforcement learning NSGA-II algorithm for multi-objective flexible job shop scheduling[J/OL]. Journal of Chongqing University,2021. (2021-05-12)[2022-01-10]. http://kns. cnki. net/kcms/detail/5 0. 1044. N. 20210511. 1913. 010. html.)
[26] Liu C, Chang C, Tseng C. Actor-Critic deep reinforcement learning for solving job shop scheduling problems[J]. IEEE Access, 2020, 8: 71752-71762.
[27] Park J, Chun J, Kim S H, *et al*. Learning to schedule job-shop problems: representation and policy learning using graph neural network and reinforcement learning[J]. International Journal of Production Research,2021,59(11):3360-3377.
[28] Roesch M, Linder C, Bruckdorfer C, *et al*. Industrial load management using multi-agent reinforcement learning for rescheduling[C]//Proc of the 2nd International Conference on Artificial Intelligence for Industries. Piscataway, NJ:IEEE Press,2019:99-102.
[29] Zhao Meng, Li Xinyu, Gao Liang, *et al*. An improved Q-learning based rescheduling method for flexible job-shops with machine failures[C]// Proc of the 15th IEEE International Conference on Automation Science and Engineering . Piscataway, NJ:IEEE Press,2019:331-337.
[30] Bär S, Turner D, Mohanty P K, *et al*. Multi agent deep Q-network approach for online job shop scheduling in flexible manufacturing[C]// Proc of International Conference on Manufacturing System and Multiple Machines. 2020:1-8.
[31] Luo Bin, Wang Sibao, Yang Bo, *et al*. An improved deep reinforcement learning approach for the dynamic job shop scheduling problem with random job arrivals[C]//Proc of the 4th International Conference on Advanced Algorithms and Control Engineering. Bristol:IOP Publishing Press,2021:1-8.
[32] Turgut Y, Bozdaag C E. Deep Q-network model for dynamic job shop scheduling problem based on discrete event simulation[C]//Proc of Winter Simulation Conference. Piscataway, NJ: IEEE Press, 2020: 1551-1559.
[33] Wang Yufang. Adaptive job shop scheduling strategy based on weighted Q-learning algorithm[J]. Journal of Intelligent Manufacturing, 2020,31(2):417-432.
[34] Luo Shu. Dynamic scheduling for flexible job shop with new job insertions by deep reinforcement learning[J]. Applied Soft Computing, 2020,91:article ID 106208.
[35] Luo Shu, Zhang Linxuan, Fan Yushun. Dynamic multi-objective scheduling for flexible job shop by deep reinforcement learning[J]. Computers & Industrial Engineering, 2021, 159: article ID 107489.
[36] Shahrabi J, Adibi M A, Mahootchi M. A reinforcement learning approach to parameter estimation in dynamic job shop scheduling[J]. Computers & Industrial Engineering,2017,110:75-82.
[37] Kardos C, Laflamme C, Gallina V, *et al*. Dynamic scheduling in a job-shop production system with reinforcement learning [J]. Procedia CIRP,2021,97:104-109.
[38] Wang Libing, Hu Xin, Wang Yin, *et al*. Dynamic job-shop scheduling in smart manufacturing using deep reinforcement learning[J]. Computer Networks,2021,190:article ID 107969.
[39] Han Wei, Guo Fang, Su Xichao. A reinforcement learning method for a hybrid flow-shop scheduling problem [J]. Algorithms, 2019, 12 (11):article ID 222.
[40] Reyna Y C F, Martínez-Jiménez Y. Adapting a reinforcement learning approach for the flow shop environment with sequence-dependent setup time[J]. Revista Cubana de Ciencias Informáticas,2017,11 (1):41-57.
[41] 张东阳,叶春明. 应用强化学习算法求解置换流水车间调度问题[J]. 计算机系统应用,2019,28(12):195-199. (Zhang Dongyang, Ye Chunming. Application of reinforcement learning algorithm to solve the permutation flow shop scheduling problem[J]. Computer Systems & Applications,2019,28(12):195-199.)
[42] 肖鹏飞,张超勇,孟磊磊,等. 基于深度强化学习的非置 换流水车间调度问题[J]. 计算机集成制造系统,2021,27(1):192-205. (Xiao Pengfei, Zhang Chaoyong, Meng Leilei, *et al*. Non-permutation flow shop scheduling problem based on deep reinforcement learning [J]. Computern Integrated Manufacturing Systems, 2021, 27 (1):192-205.)
[43] Xue Tianfang, Zeng Peng, Yu Haibin. A reinforcement learning method for multi-AGV scheduling in manufacturing [C]//Proc of IEEE International Conference on Industrial Technology. Piscataway, NJ:IEEE Press,2018:1557-1561.
[44] Arviv K, Stern H, Edan Y. Collaborative reinforcement learning for a two-robot job transfer flow-shop scheduling problem[J]. International Journal of Production Research,2016,54(4):1196-1209.
[45] Heger J, Voss T. Dynamically adjusting the *k*-values of the ATCS rule in a flexible flow shop scenario with reinforcement learning[J]. International Journal of Production Research, 2021, DOI: 10. 1080/00207543. 2021. 1943762.
[46] Reyna Y C F, Cáceres A P, Jiménez Y M, *et al*. An improvement of reinforcement learning approach for permutation of flow-shop scheduling problems [J]. Revista Ibérica de Sistemas e Tecnologias de Informação,2019(E18):257-270.
[47] 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 Ling, Pan Zixiao. Flow shop scheduling optimization based on deep reinforcement learning and iterative greedy[J]. Control and Decision, 2021, 36 (11):2609-2617.)
[48] Öztop H, Tasgetiren M F, Kandiller L, *et al*. A novel general variable neighborhood search through Q-learning for no-idle flowshop scheduling[C]//Proc of IEEE Congress on Evolutionary Computation . Piscataway, NJ:IEEE Press,2020:1-8.
[49] Zhu Jialin, Wang Huangang, Zhang Tao. A deep reinforcement learning approach to the flexible flowshop scheduling problem with makespan minimization[C]//Proc of the 9th IEEE Data Driven Control and Learning Systems Conference . Piscataway, NJ: IEEE Press, 2020: 1220-1225.
[50] Pan Ruyuan, Domg Xingye, Han Sheng. Solving permutation flowshop problem with deep reinforcement learning[C]//Proc of Prognostics and Health Management Conference. Piscataway, NJ: IEEE Press, 2020:349-353.
[51] Yang Shengluo, Xu Zhigang, Wang Junyi. Intelligent decision-making of scheduling for dynamic permutation lowshop via deep reinforcement learning[J]. Sensors,2021,21(3):article ID 1019.
[52] Yang Shengluo, Xu Zhigang. Intelligent scheduling for permutation flow shop with dynamic job arrival via deep reinforcement learning [C]//Proc of the 5th IEEE Advanced Information Technology, Electronic and Automation Control Conference . Piscataway, NJ: IEEE Press,2021:2672-2677.
[53] 刘全,翟建伟,章宗长,等. 深度强化学习综述[J]. 计算机学报,2018,41(1):1-27. (Liu Quan, Qu Jianwei, Zhang Zongchang, *et al*. Summary of deep reinforcement learning[J]. Chinese Journal of Computers,2018,41(1):1-27.)