

自动化与智能化技术

智能制造车间分布式 AGV 任务分配研究

张中伟¹, 高增恩¹, 王菁锐¹, 赵彬彬¹, 武照云^{1*}, 李鹏²

(1. 河南工业大学 机电工程学院, 河南省超硬磨料磨削装备重点实验室, 郑州 450001;

2. 郑州德力自动化物流设备制造有限公司, 郑州 452470)

摘要: **目的** 分布式自动导引车 (AGV) 系统具有可扩展性、高可靠性的优势, 是建设智能制造车间、实现车间智能物流的重要发展趋势, 研究旨在解决分布式 AGV 任务分配 (DAGVTA) 这一影响智能制造车间资源利用和生产成本的基础关键问题。**方法** 引入深度强化学习思想, 将各 AGV 视为独立的智能体, 基于多智能体强化学习方法独立深度 Q 网络 (IDQN) 算法进行求解。首先, 将 DAGVTA 问题转化为强化学习对应的部分可观测马尔可夫决策过程, 将各 AGV 观测的车间环境状态作为神经网络的输入, 通过神经网络拟合值函数输出各 AGV 的动作选择, 并以作业任务的搬运距离作为优化目标设计相应的奖励函数。在 IDQN 算法架构上训练各智能体, 环境中各智能体彼此独立, 依靠局部观测信息决策动作。**结果** 开展实验研究, 在不同问题规模场景下与基于规则的任务分配算法、基于市场机制的竞拍算法进行 DAGVTA 求解效果对比, 验证所提模型和求解方法的可行性。**结论** 基于 IDQN 架构训练后的 AGV 智能体, 在没有集中规划器的情况下具有一定的自主协同能力, 能够有效协同 AGV 完成全部搬运任务。

关键词: 智能制造车间; 分布式 AGV; 任务分配; 深度强化学习; 独立深度 Q 网络

中图分类号: TB485.3; TH187 **文献标志码:** A **文章编号:** 1001-3563(2025)07-0142-08

DOI: 10.19554/j.cnki.1001-3563.2025.07.017

Distributed AGV Task Allocation in Intelligent Manufacturing Workshops

ZHANG Zhongwei¹, GAO Zeng'en¹, WANG Jingrui¹, ZHAO Binbin¹, WU Zhaoyun^{1*}, LI Peng²

(1. Henan Key Laboratory of Superhard Abrasives and Grinding Equipment, School of Mechanical & Electrical Engineering, Henan University of Technology, Zhengzhou 450001, China; 2. Zhengzhou Deli Automation Logistics Equipment Manufacturing Co., Ltd., Zhengzhou 452470, China)

ABSTRACT: With the advantages of scalability and high reliability, the distributed automated guided vehicle (AGV) system has become a significant development trend for building intelligent manufacturing workshops and achieving intelligent logistics in the workshops. The work aims to clarify the distributed AGV task allocation (DAGVTA) in intelligent manufacturing workshops, as a fundamental key issue, that affects resource utilization and production costs. Correspondingly, deep reinforcement learning was introduced, each AGV was regarded as an independent agent, and a multi-agent reinforcement learning method independent deep Q-network (IDQN) was utilized for solution. Firstly, the DAGVTA problem was transformed into a partially observable Markov decision process related to reinforcement learning. The workshop environment states observed by each AGV were used as inputs to the neural network, and the neural

收稿日期: 2024-10-18

基金项目: 国家自然科学基金 (U1704156); 河南省科技研发计划联合基金 (242103810064); 河南省高等学校重点科研资助项目 (23A460003); 河南省超硬磨料磨削装备重点实验室开放课题资助项目 (JDKFJJ2022012)

*通信作者

network was fitted with a value function to output the action selection for each AGV. Meanwhile, the reward function was designed with the transport distance of material handling tasks as the optimization objective. Furthermore, various agents were trained on the IDQN architecture, in which each agent in the environment was independent, and took actions based on local observation information. Finally, the experimental study in scenarios with different problem scales was conducted to verify the feasibility of the proposed model and method by comparing their solution effects with rule-based task allocation algorithms and the market-based bidding algorithm. After training, the AGV agent has a certain degree of autonomous collaboration ability and can collaborate to complete all transportation tasks without a centralized planner.

KEY WORDS: intelligent manufacturing workshop; distributed AGV; task allocation; deep reinforcement learning; independent deep Q-network

在包装行业制造车间, 仓库、生产线和不同工序间存在大量纸张、塑料薄膜、纸板等原材料, 以及纸盒、模切件等半成品、成品的物料搬运活动。近年来, 自动导引车 (Automated Guided Vehicle, AGV) 凭借行动快捷、工作效率高、可控性强等优势广泛应用于包装行业制造车间的物料运输。包装行业制造车间具有动态性、复杂性、不确定性等特点, 智能制造车间调度需要高度的灵活性才能成功应对生产环境中决策规划过程的不确定性^[1]。传统制造车间中 AGV 调度广泛采用集中式调度机制, 即由中央调度器负责 AGV 系统的全部决策, AGV 基本无自主决策能力。随着 AGV 智能化水平的提升, 分布式 AGV 系统凭借可扩展性、高可靠性等优势已成为包装行业实现车间智能物流、推进智能化车间建设的重要技术支撑。

目前, 解决分布式任务分配问题的算法主要有基于规则的任务分配算法、基于市场的竞拍算法以及应用博弈论和强化学习 (Reinforcement Learning, RL) 思想的其他算法等。基于规则的任务分配算法易实现, 不需要大量计算资源, 但存在效率低、不利于全局寻优等缺点。轩华^[2]为了提高资源利用率, 利用先到先服务 (First Come First Serve, FCFS) 规则的启发式算法进行运输任务和车辆的分配; 刘奇辉等^[3]针对无人系统分布式任务分配问题, 建立考虑任务时间及路径约束下的任务分配模型, 采用基于共识的捆绑竞拍算法进行求解; De 等^[4]提出基于顺序竞拍的分散式任务分配算法, 解决分散式 AGV 控制下的智能资源管理问题。基于市场的竞拍算法应用于分布式 AGV 任务分配 (Distributed AGV Task Allocation, DAGVTA) 时, 各 AGV 通过竞标出价, 任务负载可以达到相对均衡, 且算法在时间、资源消耗方面具有优势, 但 AGV 之间互不了解对方的出价情况, 会影响决策效果。

智能制造车间 DAGVTA 问题属于序贯决策问题。RL 算法通过 Agent 与环境之间的交互对策略进行优化, 追求最大回报, 适用于解决序贯决策问题^[5]。Yuan 等^[6]针对多机器人任务分配问题, 提出基于深度强化学习 (Deep Reinforcement Learning, DRL) 的任务分配算法; Liu 等^[7]提出一种基于 DRL 的分布式

无人机蜂群任务分配算法, 解决传统 DRL 算法可扩展性不足的问题; Tang 等^[8]提出一种解决多 AGV 系统订单拣选动态调度问题的分层 RL 算法; Ma 等^[9]提出一种端到端的任务分配框架, 能够自主、智能地将任务分配给机器人车队; 李珣等^[10]针对多机器人动态任务分配问题, 提出基于博弈论的分布式自主决策框架; Liu 等^[11]研究博弈论框架中任务分配约束下的任务分配问题。以上研究表明, 结合新兴的人工智能思想和技术, 利用 AGV 的自主决策能力, 对于提高车间 AGV 任务分配的智能化、实现智能制造车间智能调度具有重要的理论和现实意义。

为适应智能制造车间 DAGVTA 更加自主灵活的需求, 构建以 AGV 搬运距离为优化目标的 DAGVTA 问题模型, 进而引入 DRL 思想, 将 DAGVTA 问题转化为马尔可夫决策过程 (Markov Decision Process, MDP), 并采用独立深度 Q 网络 (Independent Deep Q-Network, IDQN) 算法进行求解。通过开展实验研究, 将本文算法与基于规则的任务分配算法、基于市场的竞拍算法进行求解效果对比, 验证所提算法的有效性。

1 问题描述与建模

DAGVTA 可描述为随着车间生产的进行, 某一时刻系统接收到 m 个物料搬运任务, 同时车间存在 n 台空闲 AGV。空闲 AGV 需要根据自身状态自主选择合适的搬运任务, 以最短的时间或路径完成所有任务, 使系统整体物流作业效率最高。对于物料搬运任务 j ($j=1,2,\dots,m$), 其搬运起始位置 p_j 和目标位置 e_j 随任务下发而确定。任务发布时, 第 i 台 AGV ($i=1,2,\dots,n$) 在车间中的位置记为 w_i 。

为构建 DAGVTA 问题数学模型, 设计求解算法, 本文提出决策点的概念。决策点即车间同时存在物料搬运任务和可用空闲 AGV 时, 系统进入任务分配决策的时机。当 AGV 系统在 t 时刻进入决策点时, 物料搬运任务池和可用 AGV 资源池根据任务分配决策的结果同步更新状态; 当系统不再满足决策点要求, 退出决策点时, 承担新物料搬运任务的 AGV 从 t 时刻执行该任务; 随着车间生产的进行 ($t=t+1$), 系统

持续更新任务池和 AGV 资源池状态,判断何时进入下一个决策点。

在进行 DAGVTA 问题建模时,仅考虑一个任务分配决策周期,并进行以下假设:

- 1) AGV 有优先级之分,优先级高的 AGV 在物料搬运任务分配时具有优先选择权;
 - 2) AGV 为单负载,且每个物料搬运任务只能由一个 AGV 执行;
 - 3) 作业过程中不考虑 AGV 停车充电情况;
 - 4) 各 AGV 进行物料搬运作业时行驶速度相同且保持不变;
 - 5) 不考虑 AGV 路径冲突和车间障碍物的影响。
- 分布式 AGV 系统任务分配流程如图 1 所示。

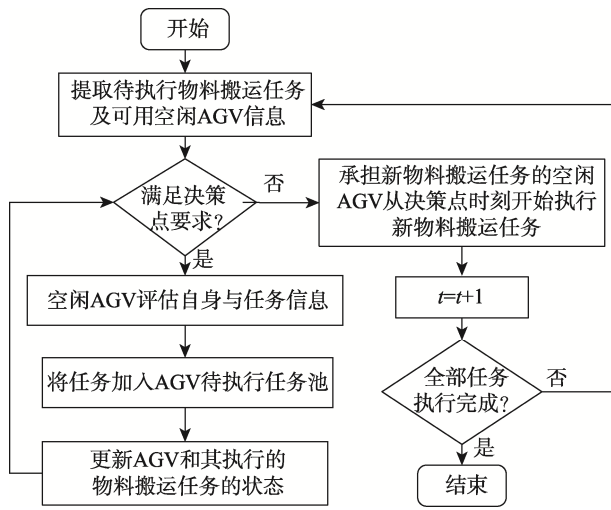


图 1 AGV 完成搬运任务流程
Fig.1 Process diagram of AGV completing handling tasks

为了评估任务分配方案的优劣,选择任务分配研究中常用的曼哈顿距离作为参考,以 AGV 完成全部任务的总搬运距离作为评价指标。为了便于量化计算 AGV 执行物料搬运任务的搬运距离,选择栅格法将车间环境建模成二维栅格地图^[12],如图 2 所示,黑色栅格表示车间中的障碍物或加工区域,白色栅格表示可通行区域。任务 j 的搬运起止位置 p_j 和 e_j 可分别表示为具体的栅格坐标,即 $p_j=(x_j, y_j)$, $e_j=(x'_j, y'_j)$,第 i 台 AGV 在任务发布时的位置 $w_i=(x_i^*, y_i^*)$,其中 x_j, x'_j 和 x_i^* 为栅格点的横坐标, y_j, y'_j 和 y_i^* 为栅格点的纵坐标。

若决策点时刻第 i 台 AGV 选择执行任务 j ,则执行任务 j 的总搬运距离 d_{ij} 可表示为式 (1)。

$$d_{ij} = d_{ij}^* + c_j \quad (1)$$

式中: d_{ij}^* 为 AGV 空载运行距离,见式 (2); c_j 为 AGV 负载搬运距离,见式 (3)。

$$d_{ij}^* = |x_i^* - x_j| + |y_i^* - y_j| \quad (2)$$

$$c_j = |x_j - x'_j| + |y_j - y'_j| \quad (3)$$

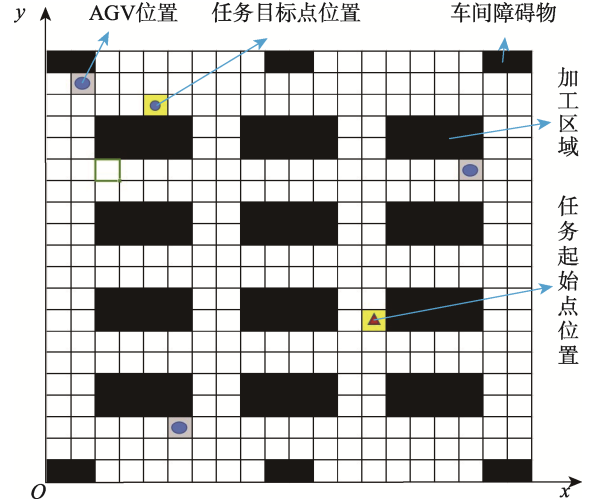


图 2 栅格地图示例
Fig.2 Example of raster map

任务 j 一旦发布, p_j 和 e_j 随之确定, c_j 可视为常量, d_{ij} 只与 d_{ij}^* 有关。因此, DAGVTA 问题优化目标实质上是 AGV 执行搬运任务的总空载运行距离。决策变量 z_{ij} 定义如下: 若第 i 台 AGV 选择执行第 j 个物料搬运任务, 则 $z_{ij}=1$, 否则 $z_{ij}=0$ 。

DAGVTA 问题的优化目标见式 (4)。

$$\min \left[\sum_{i=1}^n \sum_{j=1}^m z_{ij} d_{ij}^* \right] \quad (4)$$

约束条件见式 (5) ~ (6)。

$$\sum_{j=1}^m z_{ij} \leq 1, \forall i \quad (5)$$

$$\sum_{i=1}^n z_{ij} \leq 1, \forall j \quad (6)$$

在一个任务分配决策周期内, 式 (5) 表示每个物料搬运任务最多被一台空闲 AGV 选择执行, 式 (6) 表示每台空闲 AGV 最多选择执行一个搬运任务。

2 模型求解

RL 凭借动态适应性、探索与利用的平衡机制, 成为解决序贯决策问题的有力方法。本文采用 RL 算法, 并基于 IDQN 算法框架设计解决 DAGVTA 问题的多智能体强化学习算法, 在具体应用时重点是将任务分配问题转换成对应的 MDP。

2.1 任务分配问题的转换

使用 RL 算法解决 DAGVTA 问题需要将其转变为 $(S, A, P(s, a), R)$ 形式的 MDP, 其中 S 为状态空间, A 为动作空间, $P(s, a)$ 为状态 s 下选择动作 a 后转移至新状态 $s_'$ 的概率, R 为获得的回报。具体操作如下: 将车间的每个 AGV 看作智能体, 从车间环境中获得当前的观测状态, 通过内置策略选取一个对应当前状态的动作, 执行动作后改变环境状态并获得奖励。

2.1.1 状态空间设计

状态空间设计是智能体自主选择搬运任务的关键。由于每台 AGV 依靠局部观测状态自主决策搬运任务, 将与第 i 台 AGV 关联的智能体观测状态设计为式 (7)。

$$s_i = (o_{i1}, o_{i2}, \dots, o_{ij}) \quad (7)$$

式中: o_{ij} 为第 i 台 AGV 从当前位置到任务 j 起始位置的曼哈顿距离。如果 AGV 分配到任务 j , 其状态变为忙碌, 完成任务后状态变回空闲, 则 AGV 位置由 w_i 变为任务终点 e_j , 智能体的局部观测状态 s_i 随着 AGV 状态的变化同步更新。

2.1.2 动作空间设计

动作空间设计影响算法的搜索效率和性能。动作空间中的动作为智能体从决策点时刻的物料搬运任务池中选择一个任务, 当各智能体选择动作后, 物料搬运任务池和可用 AGV 资源池的任务与 AGV 信息进行更新, 决策变量 z_{ij} 同步改变。

2.1.3 状态转移概率

RL 算法根据是否依赖环境模型分为有模型 RL 和无模型 RL, 前者需要建立精确的环境动态模型, 计算状态转移概率 $P(s, a)$, 预测环境状态如何随动作变化; 后者在与环境的交互中学习最优策略, 无需显式构建环境的动态模型。智能制造车间环境中的任务需求可能随时间变化, AGV 之间存在竞争与合作关系, 环境动态性较强。无模型 RL 算法能够更快适应环境变化, 因此选择该算法, 避免计算状态转移概率, 通过在各状态动作对上计算累计奖励值来学习策略。

2.1.4 奖励函数设计

奖励函数的设计与优化目标紧密相关, 且影响最终的训练效果。利用 RL 算法解决任务分配问题时, 优化目标函数的取值可作为智能体的奖励值, 因此将每个智能体的奖励函数设计为式 (8)。

$$r_i = -d_{ij}^* \quad (8)$$

2.2 基于 IDQN 的算法设计

IDQN 算法是指在 RL 算法中的 Q_Learning 的基础上, 采用神经网络拟合 Q 值, 解决传统 Q_Learning 易出现的状态空间维度爆炸的问题; 设置经验池, 解决 RL 算法产生经验之间的关联性问题, 其广泛应用于单智能体路径规划、资源分配等方面。Tampuu 等^[13]应用 IDQN 算法解决 2 个智能体在合作和竞争任务场景下的问题, 为使用该算法解决分布式多智能体环境下的决策问题提供指导。在解决智能制造车间 DAGVTA 问题时, 将各 AGV 视为独立的智能体, 通过 IDQN 算法框架使每个 AGV 维持各自的神经网络。训练过程中, 智能体在决策点处从环境中获取自身观测状态 s_i , 然后将其输入各自的神经网络, 输出选择

各个动作的 Q_i 值, 最终根据 ε -greedy 策略确定选择的动作, 进入下一状态 s_{i-} , 同时获得一个奖励 r_i 。各智能体通过神经网络选择动作, 并利用 RL 算法学习不同环境状态下的动作选择策略。

2.2.1 利用与探索

DRL 算法中常用的探索策略包括 ε -greedy^[14]、玻尔兹曼探索策略^[15]和高斯策略等^[16]。探索是指随机选择动作以增加选择的多样性, 从而寻找整体最优策略; 利用是指选择当前已知最优策略的动作, 以使回报最大化。本文采用 ε -greedy 策略, 该策略提供了灵活的调节机制, 通过调节参数 ε 调整平衡探索与利用的需求。将状态 s_i 下采取行动 a_i 的概率记作 $P(s_i, a_i)$, 见式 (9) ~ (10)。

$$P(s_i, a_i) = \begin{cases} 1 - \varepsilon + \eta \cdot \text{learn_step}, & a_i = a_i^*(s_i) \\ \varepsilon - \eta \cdot \text{learn_step}, & a_i \neq a_i^*(s_i) \end{cases} \quad (9)$$

$$a_i^*(s_i) = \max_{a_i} Q_i^*(s_i, a_i, \theta_i) \quad (10)$$

式中: learn_step 为当前迭代次数; θ_i 为随机权重。

2.2.2 算法实施

RL 算法产生的数据通常为序列数据, 为消除连续数据之间的相关性, 防止网络更新过程中过拟合, IDQN 算法为智能体建立容量为 N 的经验池, 将智能体每轮的决策 (s_i, a_i, r_i, s_{i-}) 存入经验池, 训练时随机选取固定数量的样本训练。每个智能体存在 2 个相同结构, 参数更新频率不同的神经网络以保证算法的稳定性, 损失函数见式 (11)。

$$L(\theta_i) = E_{s_i} \left[r_i + \gamma \max_{a_i} \{Q(s_i, a_i, \theta_i)\} - Q(s_i, a_i, \theta_i) \right]^2 \quad (11)$$

IDQN 算法实施流程如图 3 所示。

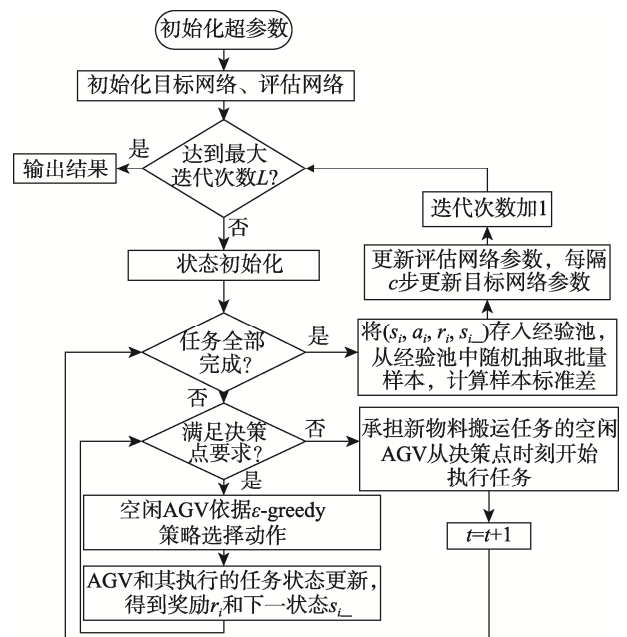


图 3 IDQN 算法流程

Fig.3 Process diagram of IDQN algorithm

1) 初始化各目标网络 $\hat{Q}_i$, 设计随机权重 θ_i ; 初始化各评估网络 Q_i , 权重参数为 $\theta_{i-} = \theta_i$ 。

2) 每轮迭代前初始化状态 $s_i = s_{i0}$ 。

3) 满足决策点要求时, 可用空闲 AGV 根据优先级进行决策, 根据当前观测状态 s_i , 以 $\varepsilon \cdot \eta \cdot \text{learn_step}$ 的概率随机选择动作 a_i , 以 $1 - \varepsilon + \eta \cdot \text{learn_step}$ 的概率选择 $a_i^*(s_i) = (s_i, a_i, \theta_i)$ 。

4) 各 AGV 和任务状态更新, 得到奖励 r_i 和下一个状态 s_{i-} 。

5) 初始状态下可用空闲 AGV 决策完成, 系统退出决策点, 承担新物料搬运任务的 AGV 开始运输, 当系统满足下一次决策点要求时, 进行下一任务选择, 直到完成所有任务。

6) 将本批次经验样本 (s_i, a_i, r_i, s_{i-}) 存入经验池, 随机抽取小批量样本计算样本目标差 $TD_{\text{error}} = r_i + \gamma \max_{a_{i-}} \{Q_i(s_{i-}, a_{i-}, \theta_{i-})\} - Q_i(s_i, a_i, \theta_i)$ 。

7) 利用损失函数 $L(\theta_i)$ 更新评估网络参数, 每隔 c 步更新目标网络参数 $\hat{Q}_i = Q_i$ 。

3 实验研究

3.1 测试实例和实验平台

算法在配置为 Intel(R) Core(TM) i9-9900K CPU@ 3.60 GHz, 32.0GB RAM Windows 11 的计算机上, 通过 Python 语言和 PyTorch 框架来实现。为验证算法的有效性, 针对系统随机生成的物料搬运任务, 与现有研究中常用的基于规则的任务分配算法、基于市场的竞拍算法进行求解结果对比。

在对比 IDQN 算法与基于规则的任务分配算法时, 分别选择常用的 FCFS 规则和 Random 规则。FCFS 规则赋予较早产生的任务更高的优先级, 即动态任务序列中到达最早的任务会被优先分配给 AGV; 在 Random 规则中, 空闲 AGV 将从任务序列中随机选择一个搬运任务, 当 AGV 执行完搬运任务, 运力释放后即有资格从物料搬运任务池中选择下一个任务执行。基于市场的竞拍算法模拟实际竞拍过程, 通过竞价的方式将任务分配给其他投标人。在对比中, 每次从任务序列中随机抽取一个任务, 各 AGV 根据当前状态进行竞标, 最终出价高者获得该任务, 直至任务全部分配完成。此外, IDQN 算法性能依赖于超参数, 本文基于参考文献[17]提供的超参数设置经验规律设置其参数, 如表 1 所示。

3.2 实验结果与分析

为便于评估任务分配方案的性能, 当将车间内部环境栅格化处理, 建模表达为二维栅格地图时, 可以定义单位栅格尺寸 (Unit Grid Cell Size, UGCS) 作为衡量车间地图尺寸和 AGV 搬运距离的基本单位,

表 1 算法参数设置
Tab.1 Algorithm parameter settings

参数	值
折扣因子 γ	0.98
贪婪策略初始 ε	0.95
贪婪策略衰减系数 η	2.375×10^{-4}
学习率 l	1×10^{-4}
记忆库容量 N	256
采样批量 b	64
目标网络参数更新频率 c	200
最大迭代次数 L	1 200

即 $\text{UGCS} = \text{车间长度方向尺寸} / \text{长度方向栅格数量} = \text{车间宽度方向尺寸} / \text{宽度方向栅格数量}$ 。相应地, 当 AGV 以单位速度 (1 m/s) 行驶时, 可以定义其移动 1 个栅格所用的时间为单位栅格移动时间 (Unit Grid Cell Movement Time, UGCMT), 并将 UGCMT 作为评价 AGV 任务执行时间的基本单位。实例 1 场景源于某包装企业制造车间, 其可抽象表达为 10 UGCS×10 UGCS 的栅格地图, 任务分配问题涉及 3 台 AGV 和 15 个物料搬运任务。训练所用的 AGV 和任务信息在每一轮迭代时随机初始化产生。采用 IDQN 算法求解该任务分配问题时, 奖励函数值随训练迭代次数的变化如图 4 所示。由于空载运行距离 (以 UGCS 为量纲) 与奖励值在数值上互为相反数, 随着算法迭代次数的增加, 空载运行距离逐渐减少, 迭代 1 000 次左右算法趋于收敛。

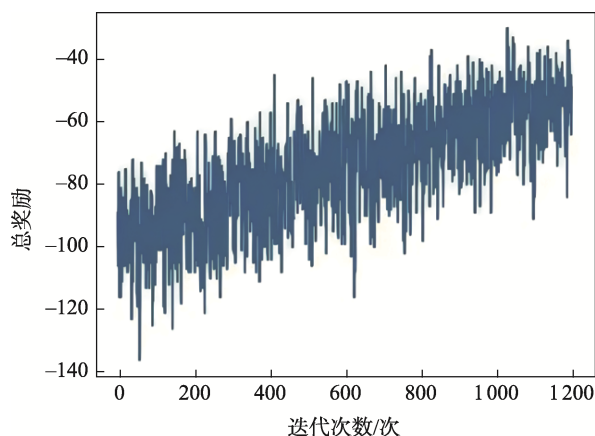


图 4 采用 IDQN 算法求解得到的奖励函数曲线
Fig.4 Reward function curve obtained by the IDQN algorithm

为了验证算法的有效性, 用实例 1 训练完成的网络进行测试, 用于测试的任务信息如表 2 所示, 3 台 AGV 的初始位置分别为 (8, 5)、(2, 8)、(8, 6)。由于研究假设 AGV 运行时速度恒定, 当设定具体的 AGV 行驶速度时, 可将物料搬运任务的分配结果通过甘特图呈现, 用来表达各 AGV 搬运任务执行顺序和相对

表 2 随机生成的 15 个物料搬运任务 Tab.2 15 randomly generated material handling tasks		
物料搬运任务编号	起始位置	目标位置
1	(9, 1)	(8, 5)
2	(7, 6)	(6, 5)
3	(1, 6)	(5, 3)
4	(6, 4)	(5, 0)
5	(7, 4)	(4, 9)
6	(0, 0)	(6, 4)
7	(9, 9)	(3, 7)
8	(4, 6)	(9, 3)
9	(2, 3)	(9, 2)
10	(9, 5)	(7, 4)
11	(2, 4)	(9, 6)
12	(3, 6)	(9, 8)
13	(7, 1)	(7, 0)
14	(9, 0)	(0, 7)
15	(5, 3)	(4, 5)

时间关系。若 AGV 均以单位速度行驶, 则实例 1 中 DAGVTA 的结果如图 5 所示, 图中 T1—T15 对应表 2 中物料搬运任务 1—15, 任务完成时间为 61 UGCMT。将任务分配结果表示为“任务索引(空载运输起始时间-负载运输起始时间-负载运输结束时间)”

的形式, AGV1 分配结果为 10 (0-1-4)→13 (4-7-8)→14 (8-10-26)→9 (26-32-40)→6 (40-51-61); AGV2 分配结果为 12 (0-3-11)→7 (11-12-20)→3 (20-23-30)→11 (30-34-43); AGV3 分配结果为 2 (0-1-3)→4 (3-4-9)→15 (9-12-15)→8 (15-16-24)→1 (24-26-31)→5 (31-33-41), 完成全部任务的空载运行距离为 44 UGCS。利用与 IDQN 算法对比的其他 3 种算法进行求解, 分配结果如表 3 所示。由图 5 和表 3 可知, IDQN 算法的任务分配结果均优于对比算法。

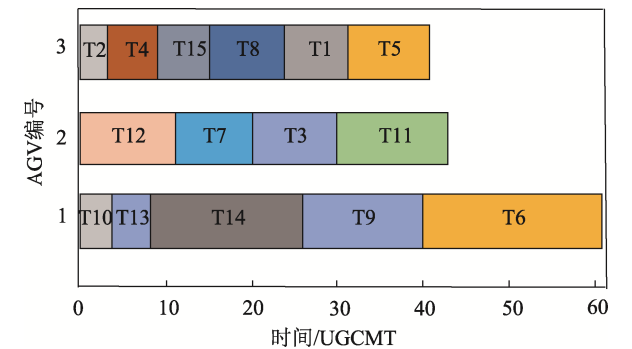


图 5 实例 1 任务分配甘特图
Fig.5 Task allocation gantt chart for Example 1

参考实例 1 场景进行拓展, 在不同规模的车间地图及不同数量的 AGV 和搬运任务场景下进行实验, 采用不同算法求得的 AGV 空载运行距离及 IDQN 算法相较于对比算法的提升率如表 4 所示, 表中数值代表迭代一定次数后, 空载运行距离在该数值附近收敛。

表 3 不同算法求解实例 1 的结果对比 Tab.3 Comparison of results for Example 1 with various algorithms				
算法	分配结果		空载搬运距离 (UGCS)	任务完成时间 (UGCMT)
基于 FCFS 规则的任务分配算法	AGV1	1 (0-5-10)→5 (10-12-20)→8 (20-23-31)→9 (31-38-46)→13 (46-49-50)→15 (50-55-58)	81	71
	AGV2	2 (0-7-9)→4 (9-10-15)→7 (15-28-36)→11 (36-40-49)→14 (49-55-71)		
	AGV3	3 (0-7-14)→6 (14-22-32)→10 (32-36-39)→12 (39-45-53)		
基于 Random 规则的任务分配算法	AGV1	8 (0-5-13)→7 (13-19-27)→13 (27-37-38)→6 (38-45-55)	98	81
	AGV2	4 (0-8-13)→9 (13-19-27)→3 (27-39-46)→15 (46-46-49)→1 (49-58-63)		
	AGV3	2 (0-1-3)→10 (3-6-9)→11 (9-14-23)→12 (23-29-37)→5 (37-43-51)→14 (51-65-81)		
竞拍算法	AGV1	10 (0-1-4)→13 (4-7-8)→5 (8-12-20)→7 (20-25-33)→9 (33-38-46)	62	70
	AGV2	8 (0-4-12)→14 (12-15-31)→11 (31-36-45)→1 (45-50-55)→3 (55-63-70)		
	AGV3	2 (0-1-3)→4 (3-4-9)→15 (9-12-15)→6 (15-24-34)→12 (34-39-47)		

表 4 不同规模下 4 种算法实验结果对比
Tab.4 Comparison of experimental results of four algorithms at different scales

实例	地图尺寸 (UGCS)	AGV 数量	任务数量	空载搬运距离 (UGCS)				IDQN 相较于对比算法的提升率/%		
				Random	FCFS	竞拍算法	IDQN	Random	FCFS	竞拍算法
1	10×10	3	15	100	100	70	60	40	40	14.2
2	20×20	8	20	260	270	210	175	32.6	35.1	16.6
3	30×30	10	40	780	800	600	550	29.4	31.2	8.3
4	50×50	20	100	3 300	3 300	2 850	2 800	15.1	15.1	1.8

由表 4 可知,在不同问题规模下, IDQN 算法均取得优于其他 3 种算法的结果。在多 AGV 系统中, AGV 之间存在任务资源的竞争, 竞拍算法中各 AGV 倾向于选择使自身利益最大化的任务, 而 IDQN 算法探索能力更好, 能够发现 AGV 之间潜在的竞争与合作关系, 注重整体利益最大化, 促使各 AGV 选择使最终整体利益最大化的任务。随着 AGV 和任务数量的增加, 状态空间呈指数增长, 分布式架构下利用 IDQN 算法能够缓解维度爆炸问题, 通过调整学习率、记忆库容量及采样批量等参数, 提升算法的扩展能力和收敛效果。尽管训练阶段 IDQN 算法迭代次数较多, 但训练后的模型可以带来更优的长期利益, 离线训练的模型可以直接投入使用, 避免大量实时计算消耗。在实例 4 场景下, IDQN 算法随着迭代次数的增加, 奖励曲线逐步增加并趋于收敛 (图 6), 这表明 IDQN 算法具有较好的泛化能力。

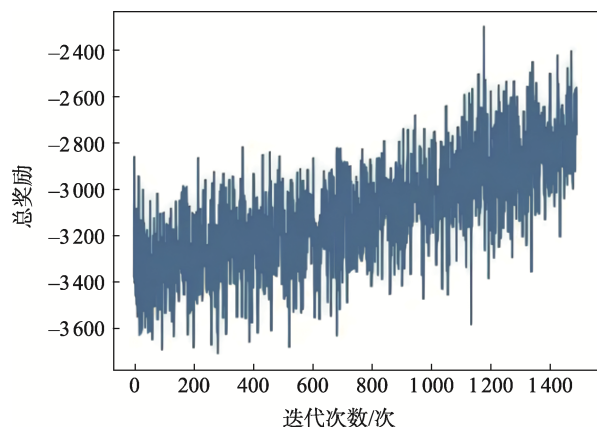


图 6 采用 IDQN 算法求解实例 4 得到的奖励函数曲线

Fig.6 Reward function curve obtained by the IDQN algorithm from Example 4

4 结语

在当前智能制造车间应用分布式 AGV 系统, AGV 调度策略智能化程度亟须提高的背景下, 以 AGV 搬运距离为优化目标构建 DAGVTA 问题模型, 进而引入 DRL 思想, 将 DAGVTA 问题转化为 MDP,

利用 IDQN 算法求解。所采用的算法适用于以包装行业智能制造车间为代表的多辆 AGV、物料搬运任务起止位置灵活变化的场景, 在不同问题规模下, 同基于 FCFS 规则、基于 Random 规则的算法和基于市场机制的竞拍算法相比, 求解效果提升 1.8%~40%, 验证了所提算法的可行性和有效性。未来研究将综合考虑能耗、设备利用率等多个优化目标, 并集成考虑 AGV 避障与路径规划问题, 提升 AGV 系统的整体性能, 从而更好地满足智能制造车间的需求。

参考文献:

- [1] SERRANO-RUIZ J C, MULA J, POLER R. Job Shop Smart Manufacturing Scheduling by Deep Reinforcement Learning[J]. Journal of Industrial Information Integration, 2024, 38: 100582.
- [2] 轩华. 基于 FCFS 策略的带时间窗车队调度问题研究[J]. 交通运输系统工程与信息, 2013, 13(6): 140-146.
XUAN H. Fleet Scheduling Problem with Time Windows Based on FCFS Strategy[J]. Journal of Transportation Systems Engineering and Information Technology, 2013, 13(6): 140-146.
- [3] 刘奇辉, 刘海颖, 黄魁华, 等. 基于 CBBA 的多域作战无人系统任务规划方法研究[J]. 火力与指挥控制, 2024, 49(2): 87-94.
LIU Q H, LIU H Y, HUANG K H, et al. Research on Mission Planning Method of Multi-Domain Operation Unmanned Systems Based on CBBA[J]. Fire Control & Command Control, 2024, 49(2): 87-94.
- [4] DE R M, PISSOORT D, HOLVOET T, et al. Decentral Task Allocation for Industrial AGV-Systems with Resource Constraints[J]. Journal of Manufacturing Systems, 2021, 59: 310-319.
- [5] 李茹杨, 彭慧民, 李仁刚, 等. 强化学习算法与应用综述[J]. 计算机系统应用, 2020, 29(12): 13-25.
LI R Y, PENG H M, LI R G, et al. Overview on Algorithms and Applications for Reinforcement Learning[J]. Computer Systems & Applications, 2020, 29(12): 13-25.

- [6] YUAN R P, DOU J T, LI J T, et al. Multi-Robot Task Allocation in E-Commerce RMFS Based on Deep Reinforcement Learning[J]. Mathematical Biosciences and Engineering, 2023, 20(2): 1903-1918.
- [7] LIU B, WANG S L, LI Q H, et al. Task Assignment of UAV Swarms Based on Deep Reinforcement Learning[J]. Drones, 2023, 7(5): 297.
- [8] TANG H L, WANG A Q, XUE F, et al. A Novel Hierarchical Soft Actor-Critic Algorithm for Multi-Logistics Robots Task Allocation[J]. IEEE Access, 2021, 9: 42568-42582.
- [9] MA S, RUAN J Q, DU Y L, et al. An End-to-End Deep Reinforcement Learning Based Modular Task Allocation Framework for Autonomous Mobile Systems[J]. IEEE Transactions on Automation Science and Engineering, 2024, 22: 1519-1533.
- [10] 李珣, 南恺恺, 赵征凡, 等. 多智能体博弈的纺织车间搬运机器人任务分配[J]. 纺织学报, 2020, 41(7): 78-87.
- LI X, NAN K K, ZHAO Z F, et al. Task Allocation of Handling Robot in Textile Workshop Based on Multi-Agent Game Theory[J]. Journal of Textile Research, 2020, 41(7): 78-87.
- [11] LIU C X, LU K H, CHEN X J, et al. Game-Theoretical Approach for Task Allocation Problems with Constraints[J]. Applied Mathematics and Computation, 2023, 458: 128251.
- [12] 朱敏, 胡若海, 卞京. 基于改进蚁群算法的移动机器人路径规划[J]. 现代制造工程, 2024(3): 38-44.
- ZHU M, HU R H, BIAN J. Path Planning for Mobile Robots Based on Improved Ant Colony Algorithm[J]. Modern Manufacturing Engineering, 2024(3): 38-44.
- [13] TAMPUU A, MATHISEN T, KODELJA D, et al. Multi-agent Cooperation and Competition with Deep Reinforcement Learning[J]. PLoS One, 2017, 12(4): e0172395.
- [14] KHLIF N, NAHLA K, SAFYA B. Reinforcement Learning with Modified Exploration Strategy for Mobile Robot Path Planning[J]. Robotica, 2023, 41(9): 2688-2702.
- [15] 董瑶, 葛莹莹, 郭鸿湧, 等. 基于深度强化学习的移动机器人路径规划[J]. 计算机工程与应用, 2019, 55(13): 15-19.
- DONG Y, GE Y Y, GUO H Y, et al. Path Planning for Mobile Robot Based on Deep Reinforcement Learning[J]. Computer Engineering and Applications, 2019, 55(13): 15-19.
- [16] PENG S L, FENG Q M. Reinforcement Learning with Gaussian Processes for Condition-Based Maintenance[J]. Computers & Industrial Engineering, 2021, 158: 107321.
- [17] JIANG L, HUANG H Y, DING Z H. Path Planning for Intelligent Robots Based on Deep Q-Learning with Experience Replay and Heuristic Knowledge[J]. IEEE/CAA Journal of Automatica Sinica, 2020, 7(4): 1179-1189.