现代排序论与作业调度优化专题

DOI:10.11721/cqnuj20260203

基于 Actor-Critic 的集装箱码头箱位分配与场桥调度协同优化*

刘战龙[1]，王　祎[2]，孙晓驰[3]，李一峰[1]，张新功[1]
(1. 重庆师范大学 重庆国家应用数学中心，重庆 401331；2. 重庆港股份有限公司，重庆 400025；
3. 重庆果园集装箱码头有限公司，重庆 401133)

摘要:集装箱码头作为国际贸易的关键节点,它的作业效率直接影响船舶周转时长与物流成本。箱位分配与场桥调度是集装箱码头作业中的 2 个关键环节,二者相互耦合、相互影响。因此,提出一种基于深度强化学习 Actor-Critic 框架方法研究箱位分配与场桥调度的协同优化问题,旨在实现箱位分配与场桥调度的协同决策。一方面,建立了考虑场桥作业均衡、场桥非装卸时间、场桥作业时间、内集卡等待时间和翻箱量因素的混合整数规划模型;另一方面,通过构建集装箱堆场状态空间、场桥动作空间及多目标奖励函数模型,提出了基于深度强化学习 Actor-Critic 框架的求解算法,通过不同规模的算例对比分析,验证了所提算法与传统遗传算法在求解上具有优越性。与已有研究结果相比,在保持场桥作业均衡的前提下,新算法得到的结果能有效降低场桥作业的完成时间和非装卸时间、缩短自动导引车的等待时长以及降低翻箱率。

关键词:箱位分配;场桥调度;深度强化学习;Actor-Critic;混合整数规划模型;集装箱码头
中图分类号:O224　　**文献标志码**:A　　**文章编号**:1672-6693(2026)02-0026-16

近些年来,随着“一带一路”倡议的提出和推广,中国国际贸易规模持续扩大。根据交通运输部的数据,2023 年全国货物吞吐量累计达 1 697 326 万 t,同比增速为 8.2%;集装箱吞吐量累计达 31 034 万标准箱,增长率为 4.9%;2024 年全国货物吞吐量累计达 1 759 501 万 t,同比增速为 3.7%,集装箱吞吐量累计达 33 200 万标准箱,增长率为 7.0%。面对日益增长的集装箱规模,如何提升集装箱码头的作业效率成为码头的重要任务。然而,随着船舶大型化与货物吞吐量激增,提升集装箱码头运营效率面临着堆场空间紧张、设备调度复杂等挑战。因此,优化集装箱码头的箱位分配和场桥调度具有重要研究意义。

在集装箱码头的箱位分配问题上,刘瑞麟[1]以场桥的行车成本和翻箱成本之和最低为目标,建立堆场箱位分配模型,运用遗传算法(genetic algorithm,GA)对模型进行求解,得到具体箱位调度分配方案。余明珠等人[2]将船舶靠泊随机性和出口集装箱集港顺序综合考虑到整数规划模型中,设计了基于仿真优化思想的启发式算法求得集装箱贝位分配量,并提出“长途箱压短途箱”的贝内具体落位策略。Yang 等人[3]针对自动化集装箱码头箱区空间不足的问题,建立了混合整数二次规划模型,并设计了改进粒子群算法优化箱位分配问题。郑云峰等人[4]考虑码头到港船舶目的地、时间、箱重等因素,基于确定任务视角,将贝位选择和贝内排箱过程看作一个整体,设计了一种在整个区域内搜索排箱位置的 GA 来对模型进行求解,得到具体箱位调度分配方案。

在集装箱码头的场桥调度问题上,He 等人[5]考虑了集装箱船舶到港时间和任务量的不确定性,构建了混合整数规划模型,并提出了一个基于 GA 的三阶段框架来求解该模型。Quintero-Araújo 等人[6]基于 GA 研究了多目标场桥调度问题,考虑了多个优化目标,如降低作业成本、缩短作业时间以及提高码头作业的流转效率。Chang 等人[7]研究了单箱区双场桥协同调度问题,构建了以任务延迟数最小化为目标的整数规划模型,设计了基于滚动时域框架的启发式优化算法来求解。郑红星等人[8]考虑到作业过程中多个场桥间需保持安全距离且不可互相跨越、实际作业中积累的经验性翻箱规则等约束,构建了以最小化场桥总行走时间为目标的线性规划模型。

针对箱位分配与场桥调度的协同优化,学者已进行了不少研究[9-12],但仍存在一些问题亟待解决。范厚明等

* **收稿日期**:2025-09-29　**修回日期**:2025-12-23　**网络出版时间**:2026-05-14T08:36
资助项目:国家自然科学基金青年基金项目(No. 62403088);重庆市教育委员会科学技术研究重点项目(No. KJZD-K202400503,No. KJZD-K202500501);重庆市自然科学基金创新发展联合基金项目(No. CSTB2023NSCQ-LZX0142)
第一作者简介:刘战龙,男,研究方向为多目标优化,E-mail:2092005521@qq.com;**通信作者简介**:李一峰,男,副教授,博士,E-mail:liyifeng@cqnu.edu.cn
网络出版地址:https://link.cnki.net/urlid/50.1165.N.20260513.1821.002

人[9]提出分区域平衡策略，以均衡各场桥作业任务量和减少场桥的非装卸时间为目标，建立了混合整数规划模型解决箱位分配和场桥调度问题；但文中箱位分配只精确到贝，没有精确到具体箱位（贝、栈、层）。马晓斌[10]以场桥完工时间最短构建了送箱集卡失约下多箱区箱位分配及场桥调度协同优化模型，并设计了不同规模的算例说明设计的算法及优化模型对研究问题的有效性；但文中只针对最短完工时间进行优化，无法直接反映不同目标之间的冲突或协同关系。董良才等人[11]针对出口箱的箱位分配和场桥调度问题，建立了以减少场桥作业时间、内集卡等待时间和箱区翻箱量为目标的混合整数规划模型，并设计了一种基于深度强化学习的模型求解方法；但未考虑优化场桥的非装卸时间（场桥跑位时间和场桥空闲时间），而场桥跑位时间和场桥空闲时间是评价场桥作业效率的重要参考指标。

还有一些学者借助强化学习（reinforcement learning，RL）来解决这类复杂调度问题。该理论具有较强的环境交互能力和学习能力，为解决此类问题提供了新思路。王小寒等人[13]基于共享堆场，考虑进出口双向货流，以海铁联运集装箱在港中转成本最小化为目标，构建箱区分配及箱区转场问题的非线性整数规划模型，并设计了基于 Q 学习的协同进化 GA 求解问题。王无印等人[14]针对场桥调度问题的复杂时空耦合特性和高度动态性，以最小化自动导引车（automated guided vehicle，AGA）和外集卡的等待时间为优化目标构建数学规划模型，并提出了一种新颖的深度强化学习方法进行求解。董良才等人[11]提出基于 Actor-Critic 的堆场策划与场桥调度方案，仿真实验表明，基于 Actor-Critic 的求解算法在小规模问题上能够快速准确求解，在大规模问题上，相较于 GA 展现了显著的优越性。Liu 等人[15]提出了一种基于深度强化学习的方法，以解决场桥调度与箱位分配的协同优化问题，并通过设计多智能体系统，实现了动态环境下的实时决策，明显提高了码头作业效率。

综上所述，箱位分配和场桥调度是集装箱码头作业中的 2 个重要环节，二者相互耦合，相互影响。协同优化箱位分配和场桥调度不仅能降低翻箱率、均衡场桥作业量，还能缩短场桥非装卸时间和场桥作业时间。关于集装箱码头箱位分配和场桥调度的研究已有不少，总体上可分为 2 类：一类是仅考虑箱位分配或场桥调度问题，没有将二者进行协同研究；另一类是考虑箱位分配和场桥调度的协同优化问题，但目标函数并没有将影响箱位分配和场桥调度的诸多因素考虑进去，或研究结果中箱位分配只分配到贝位而没有精准到具体箱位（贝、列、层）。基于此，本文综合考虑箱位分配和场桥调度影响的多种因素：翻箱量、AGV 等待时间、场桥作业均衡、场桥非装卸时间和场桥作业时间，建立箱位分配与场桥调度的协同优化模型。同时，借助深度强化学习理论框架，设计基于 Actor-Critic 的求解算法。最后，与已有文献方法进行对比，以说明本文所得结果的优越性。

1　问题描述与模型构建

1.1　场桥配置的协同调度机制研究

从资源规划的角度，集装箱码头堆场内的每个箱区作为基础作业单元，内部通常部署有 2 台及以上场桥同时协同作业。然而，多场桥协同作业模式在实际运营中往往面临三重技术挑战：首先是设备负载不均衡引发的长距离大范围位移问题；其次是无效能耗增加导致的设备利用率下降现象；再者是作业空间重叠造成的协同效率制约，特别是相邻设备必须维持最小安全间距以避免物理干涉。判断场桥之间是否存在冲突在于观察场桥作业的执行轨迹在坐标系中是否存在交点。例如，图 1 中的 2 台场桥无作业冲突，而图 2 中 2 台场桥则存在作业冲突。

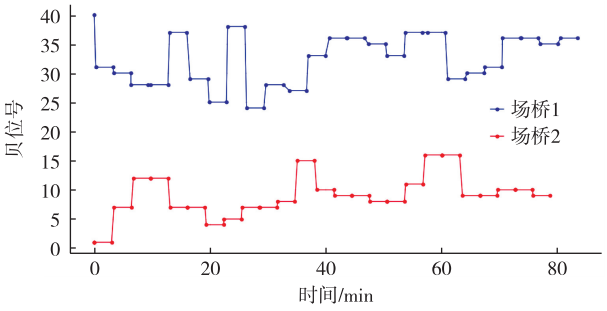


图 1　场桥作业无冲突

Fig. 1　No conflict in yard crane operation

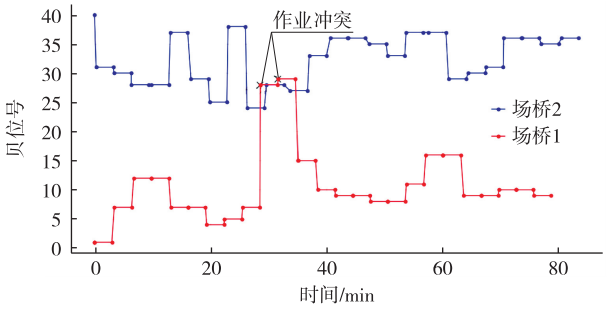


图 2　场桥作业冲突

Fig. 2　Yard crane operation conflict

基于集装箱码头堆场作业的安全性与效率要求,本文所运用的分区均衡调度策略,在箱区划分上遵循的逻辑是:以场桥的配置规模为基准,动态地将箱区划分为对应数量的作业区域。确保各区块作业负载均衡且满足设备安全间距约束。图3是分区域作业的一种情况,有3台场桥,场桥1、场桥2和场桥3分别负责区域A、区域B和区域C中的作业任务。

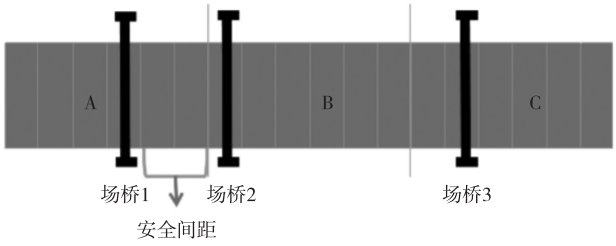


图3 分区平衡策划示意图

Fig. 3 Schematic diagram of zoning balance planning

1.2 堆场内部箱位分配问题

集装箱码头堆场作为货物装卸与暂存的核心作业区,空间资源配置效率直接影响着整体运营效能。科学的箱位分配策略对提升码头吞吐能力具有关键作用。在集装箱堆场管理中,箱位采用三维坐标系进行精确定位,即以贝、栈、层为3个坐标轴,箱位的坐标记为$Q\{\gamma,\varepsilon,\theta\}$。从作业效率的角度来看,箱位分配问题本质上可以转化为场桥移动成本和翻箱操作成本的最小化问题。这一转化基于以下2个关键作业特征:首先,当相邻作业的集装箱位于不同贝位时,场桥必须进行跨贝移动,这种移动既消耗能源,又增加作业时间;其次,在提取被堆叠的下层集装箱时,必须先将上层的阻碍集装箱转移至临时位置,这种翻箱操作不仅增加了作业复杂度,还可能引发额外的设备损耗和作业延误。

需要说明的是,翻箱操作的触发与集装箱的优先级体系密切相关。集装箱翻箱问题(container relocation problem,CRP)在相关文献中也称为块重定位问题(block relocation problem,BRP)[16],解决此问题的核心在于通过优先级评估来决定是否需要翻箱。本文优先级系统主要考虑2个关键参数:卸货港和重量。首先,根据卸货港的远近,依次界定箱区位置。其次,根据国际标准,本文根据重量将出口箱划分成4个等级,并执行"重压轻"的堆存规则。这2个参数共同构成两位数的优先级编码,例如"24"表示卸货港优先级为2,重量优先级为4。每个集装箱都具有唯一的优先级编码,对应着特定的目标堆存位置。翻箱具体决策逻辑如下:优先比较卸货港优先级,距离较远者优先;当卸货港相同时,则比较重量等级。当目标集装箱上方存在优先级更高的集装箱时,就必须进行翻箱操作。如h层为$\pi_1=14$,$h-1$层为$\pi_2=23$,则$\pi_1<\pi_2$,产生1次翻箱。如h层为$\pi_1=23$,$h-1$层为$\pi_2=24$,则$\pi_1<\pi_2$,也产生1次翻箱。图4给出了翻箱操作示意图。

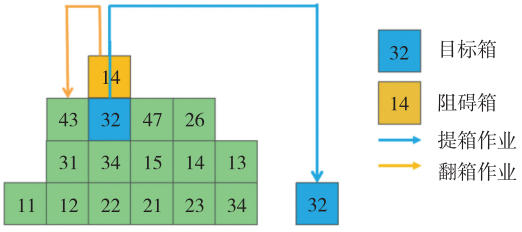


图4 提箱作业和翻箱作业

Fig. 4 Container pick-up operation and container reshuffling operation

1.3 问题假设

1)本文假设调度决策基于以下已知信息:箱区待堆存集装箱的数量与物理属性、关联AGV的预计抵达时间以及根据岸桥调度计划反推得出的近似装船顺序。2)设定集装箱堆场场桥不能跨箱区作业。3)每个箱区中场桥设备的移动速度、作业速度已知,且均保持匀速运动。4)在模型构建与运算的初始阶段,堆场内所有集装箱的初始堆存布局(包括各箱位的占用情况、集装箱的具体位置等)已明确可知,同时场桥设备在作业启动前的初始停放位置也已预先确定,为后续作业规划提供固定的初始参数。5)本文以20GP标准集装箱作为基准规格,且严格遵循"1个箱位对应1个集装箱"的堆存规则,不存在箱位共享或不同规格集装箱混存的情况,以简化堆存

空间规划逻辑。6）本文模型的研究范围聚焦于普通集装箱的堆存作业，暂不将冷藏集装箱、危险货物集装箱等具有特殊存储要求或安全管控需求的特种集装箱纳入考量。7）所有待处理任务箱，均需按照规划流程堆存至指定堆场区域，不存在直提箱。

1.4　参数设置

本文定义的主要参数、变量符号及相关含义见表 1。

表 1　符号说明

Tab. 1　Explanation of symbol parameters

类型	符号	意义及说明	类型	符号	意义及说明
参数	Ω	需要处理的出口集装箱相关作业任务的数量总和	中间变量	Θ_λ	场桥 λ 完成的作业总量
	p,q	任务集合中的某 2 个任务 p,q（$p,q\in\Omega$）是作业前后关系		β_p	任务 p 的目标贝位
	β	箱区内贝位 γ 的数量		δ_p	任务 p 的目标栈位
	δ	箱区内栈 ε 的数量		t_λ^{01}	场桥 λ 由初始位置首任务位的初始空载耗时
	η	单个贝内可堆存的集装箱 θ 的总数量，即 $\theta=1,2,\cdots,\eta$		t_λ^{pq}	场桥 λ 由任务 p 位置移动到紧后任务 q 位置的在连续任务位间的空载行驶耗时
	κ	箱区配置的场桥 λ 总数		$t_{\lambda p}$	场桥 λ 执行任务 p 的作业耗时
	$\{\gamma,\varepsilon,\theta\}$	第 γ 贝 ε 栈 θ 层的箱位		$T_{\lambda p}$	场桥 λ 第 p 次任务的完成时刻
	π_p	出口箱 p 的整体优先级		$T_{\lambda\Theta_\lambda}$	场桥 λ 最后一项任务的完成时刻
	Ψ	足够大的正数		$T_{\lambda p}^{\mathrm{AGVa}}$	场桥 λ 第 p 次任务对应的 AGV 到达堆场的时刻
	Ω_0	箱区内初始堆存的出口箱箱位集合		$t_{\lambda p}^{\mathrm{wait}}$	场桥 λ 完成第 p 次任务时，对应 AGV 的等待时间
	β_λ^0	场桥 λ 初始所处的贝位		$\overline{\Theta}$	场桥平均作业任务量
	δ_λ^0	场桥 λ 初始所处的栈		$T_{\lambda p}^{\mathrm{arrive}}$	场桥 λ 到达任务 p 作业位置的时刻
	t_β	场桥跨越一个贝位所消耗的时间		$t_{\lambda p}^{\mathrm{NA}}$	场桥 λ 的空闲时间
	t_δ	场桥跨越一栈所消耗的时间	决策变量	$\Phi_{\gamma\varepsilon\theta}^p$	$\begin{cases}1,\text{任务 } p \text{ 分配给箱位 } \gamma\varepsilon\theta\\0,\text{任务 } p \text{ 未分配给箱位 } \gamma\varepsilon\theta\end{cases}$
	t_c	场桥提取或放下一个出口箱的时间		$\Gamma_{\lambda p}$	$\begin{cases}1,\text{任务 } p \text{ 分配给场桥 } \lambda\\0,\text{任务 } p \text{ 未分配给场桥 } \lambda\end{cases}$
	β_λ^p	场桥 λ 执行任务 p 所处的贝位			
	δ_λ^p	场桥 λ 执行任务 p 所处的列			
	l	单位贝位的长度			
	l_{safe}	相邻场桥之间的安全距离			

1.5　箱位分配和场桥调度协同优化模型

根据以上假设条件和参数变量设计构建的箱位分配和场桥调度问题的数学模型如下：

$$F_1:\min\ \mathcal{Z}=\max_{1\leqslant\lambda\leqslant\kappa}T_{\lambda\Theta_\lambda},\tag{1}$$

$$F_2:\min\ \mathcal{W}=\sum_{\lambda=1}^{\kappa}\sum_{p=1}^{\Theta_\lambda}t_{\lambda p}^{\mathrm{wait}},\tag{2}$$

$$F_3:\min\ \mathcal{U}=\sum_{\gamma=1}^{\beta}\sum_{\varepsilon=1}^{\delta}\sum_{\theta=1}^{\eta}\mathcal{R}_{\gamma\varepsilon\theta},\tag{3}$$

$$F_4:\min\ \mathcal{P}=\sum_{\lambda=1}^{\kappa}(\Theta_\lambda-\overline{\Theta})^2,\tag{4}$$

$$F_5:\min\ \mathcal{X}=\sum_{\lambda=1}^{\kappa}\sum_{p=1}^{\Theta_\lambda}t_{\lambda p}^{\mathrm{NA}}+\sum_{\lambda=1}^{\kappa}\sum_{p=1}^{\Theta_\lambda}t_\lambda^{pq}。\tag{5}$$

其中：式(1)表示最小化所有场桥装船任务的最大完成时间。式(2)表示最小化 AGV 在堆场处于等待状态的总

时间。式(3)表示最小化箱区总翻箱量,$\mathcal{R}_{\gamma\varepsilon\theta}$ 表示贝 γ、栈 ε、层 θ 处的翻箱次数,优先级低的箱堆放在优先级高的箱上方时触发翻箱。式(4)中 $\mathcal{P}$ 表示各场桥作业任务量的方差。式(5)中 $\mathcal{X}$ 表示为场桥的非装卸时间,包括场桥空闲时间和跑位时间。

约束条件:

$$\sum_{p=1}^{\Omega}\Phi_{\gamma\varepsilon\theta}^{p}\leqslant 1,1\leqslant\gamma\leqslant\beta,1\leqslant\varepsilon\leqslant\delta,1\leqslant\theta\leqslant\eta; \tag{6}$$

$$\sum_{\gamma=1}^{\beta}\sum_{\varepsilon=1}^{\delta}\sum_{\theta=1}^{\eta}\Phi_{\gamma\varepsilon\theta}^{p}=1,1\leqslant p\leqslant\Omega; \tag{7}$$

$$\sum_{p=1}^{\Omega}\Phi_{\gamma\varepsilon\theta}^{p}\leqslant\sum_{p=1}^{\Omega}\Phi_{\gamma\varepsilon(\theta-1)}^{p},1\leqslant\gamma\leqslant\beta,1\leqslant\varepsilon\leqslant\delta,2\leqslant\theta\leqslant\eta; \tag{8}$$

$$\sum_{p=1}^{\Omega}\Phi_{\gamma\varepsilon\theta}^{p}=0,\{\gamma,\varepsilon,\theta\}\in\Omega_{0}; \tag{9}$$

$$\sum_{\lambda=1}^{\kappa}\Gamma_{\lambda p}=1,1\leqslant p\leqslant\Omega; \tag{10}$$

$$\sum_{p=1}^{\Omega}\sum_{\varepsilon=1}^{\delta}\sum_{\theta=1}^{\eta}\Phi_{\gamma\varepsilon\theta}^{p}\leqslant 22,1\leqslant\gamma\leqslant\beta; \tag{11}$$

$$\sum_{p=1}^{\Omega}\pi_{p}\Phi_{\gamma\varepsilon\theta}^{p}+\Big(1-\sum_{p=1}^{\Omega}\Phi_{\gamma\varepsilon\theta}^{p}\Big)\Psi-\sum_{p=1}^{\Omega}\pi_{p}\Phi_{\gamma\varepsilon(\theta-1)}^{p}+\mathcal{R}_{\gamma\varepsilon\theta}\Psi\geqslant 0,1\leqslant\gamma\leqslant\beta,1\leqslant\varepsilon\leqslant\delta,2\leqslant\theta\leqslant\eta; \tag{12}$$

$$T_{\lambda p}\geqslant T_{\lambda q}+t_{\beta}(\beta_{p}-\beta_{q}+l_{\text{safe}})-\Psi(\Gamma_{\lambda p}+\Gamma_{\lambda q}),1\leqslant p,q\leqslant\Omega,1\leqslant\lambda\leqslant\kappa,\lambda\neq\lambda'; \tag{13}$$

$$\beta_{p}=\sum_{\gamma=1}^{\beta}\gamma\Phi_{\gamma\varepsilon\theta}^{p},1\leqslant\gamma\leqslant\beta,1\leqslant\varepsilon\leqslant\delta,1\leqslant\theta\leqslant\eta; \tag{14}$$

$$\delta_{p}=\sum_{\varepsilon=1}^{\delta}\varepsilon\Phi_{\gamma\varepsilon\theta}^{p},1\leqslant\gamma\leqslant\beta,1\leqslant\varepsilon\leqslant\delta,1\leqslant\theta\leqslant\eta; \tag{15}$$

$$\sum_{\lambda=1}^{\kappa}\Theta_{\lambda}=\Omega; \tag{16}$$

$$t_{\lambda}^{pq}=t_{\beta}(\beta_{\lambda}^{p}-\beta_{\lambda}^{q})+t_{\delta}(\delta_{\lambda}^{p}-\delta_{\lambda}^{q}),1\leqslant\lambda\leqslant\kappa,1\leqslant p,q\leqslant\Theta_{\lambda}; \tag{17}$$

$$t_{\lambda p}=t_{c}+t_{\lambda}^{pq},1\leqslant\lambda\leqslant\kappa,1\leqslant p,q\leqslant\Theta_{\lambda}; \tag{18}$$

$$T_{\lambda q}=T_{\lambda p}+t_{\lambda}^{pq}+t_{c},1\leqslant\lambda\leqslant\kappa,1\leqslant p,q\leqslant\Theta_{\lambda}; \tag{19}$$

$$t_{\lambda p}^{\text{wait}}=T_{\lambda p}-T_{\lambda p}^{\text{arrive}},1\leqslant\lambda\leqslant\kappa,1\leqslant p\leqslant\Theta_{\lambda}; \tag{20}$$

$$T_{\lambda p}^{\text{arrive}}=T_{\lambda q}+t_{\lambda}^{qp},1\leqslant\lambda\leqslant\kappa,1\leqslant p\leqslant\Theta_{\lambda}; \tag{21}$$

$$t_{\lambda p}^{\text{NA}}=\max\{T_{\lambda p}^{\text{arrive}}-T_{\lambda p}^{\text{AGVa}},0\},1\leqslant\lambda\leqslant\kappa,1\leqslant p\leqslant\Theta_{\lambda}; \tag{22}$$

$$\overline{\Theta}=\frac{1}{\kappa}\sum_{\lambda=1}^{\kappa}\Theta_{\lambda}; \tag{23}$$

$$\Theta_{\lambda}=\sum_{p=1}^{\Omega}\Gamma_{\lambda p}。 \tag{24}$$

其中:式(6)表示每个箱位上至多容纳1个出口箱。式(7)表示每个出口箱必须分配至唯一箱位。式(8)表示上层箱位的堆存数量不得超过下层,避免悬空堆存(即必须先存下层,再存上层)。式(9)表示针对已预先放置初始出口集装箱的特定箱位,模型通过约束明确禁止进行二次分配或调整,确保初始堆存状态的稳定性,避免因重新分配导致前期规划失效或产生额外作业成本。式(10)表示模型要求堆场内所有作业任务(如堆箱、取箱、翻箱等)均需遵循"单任务-单设备"的分配逻辑,即每个任务必须由1台场桥独立承接,且同一任务不得由多台场桥共同处理,以此保障作业流程的清晰性与设备负载统计的准确性。式(11)表示了堆场内每个贝位的物理结构固定为"6列×4层",理论上包含24个独立箱位;但考虑到装船作业中可能产生的翻箱需求(需临时腾挪空间),模型强制为每个贝位预留2个专用翻箱空间,因此每个贝位实际可用于常规堆存的箱位数量限定为22个。式(12)表示模型通过优先级对比明确翻箱作业的启动逻辑,即当待操作集装箱(任务箱)的整体优先级低于它正下方堆叠的出口集装箱优先级时,即触发一次翻箱作业。进一步细化来看,对于同一贝位、同一栈构成的"垂直栈":若该栈内出口箱堆存层数超过2层,需满足上述优先级约束 $\sum_{p=1}^{\Omega}\pi_{p}\Phi_{\gamma\varepsilon\theta}^{p}\geqslant\sum_{p=1}^{\Omega}\pi_{p}\Phi_{\gamma\varepsilon(\theta-1)}^{p}+\mathcal{R}_{\gamma\varepsilon\theta}\Psi$,以规避不必要的

翻箱；若堆放层数不超过 2 层，则需满足约束 $\Psi \geqslant \sum_{p=1}^{\Omega} \pi_p \Phi^{p}_{\gamma\varepsilon(\theta-1)} + \mathcal{R}_{\gamma\varepsilon\theta}\Psi$。式(13)表示为避免多台场桥在同一贝位区域发生作业冲突，模型设定空间冲突处理机制，即当某一贝位已被一台场桥占用时，其他需进入该贝位作业的场桥，必须等待当前占用场桥完全离开该贝位后，方可启动作业，以保障场桥运行的安全性与有序性。式(14)和(15)分别表示模型通过 2 项约束分别明确出口箱的堆存定位逻辑：其中一项约束规定每个出口箱需被分配至唯一的目标贝位，另一项约束则进一步确定该出口箱在目标贝位内的具体栈位，二者相结合实现出口箱堆存位置的精准锁定。式(16)表示模型要求所有场桥需处理的作业任务总量，与堆场内待堆存出口箱对应的总作业需求完全相等，确保作业任务分配无遗漏、无冗余，实现场桥资源与作业需求的精准匹配。式(17)表示模型针对场桥执行单个作业任务的场景，计算出场桥在不同位置间移动所需的跑位时间。式(18)表示模型明确场桥完成单个作业任务的总时间构成，不仅包含前述的移动时间，还涵盖了集装箱的抓取、提升、放置等操作时间，通过整合各环节耗时，得到单个任务的完整执行周期。式(19)表示后一个任务的启动时间节点，确保前后任务衔接无间隙、无重叠，保障场桥作业流程的连续性与高效性。式(20)表示 AGV 在堆场中的等待总时间。式(21)表示场桥作业任务 p 的到达时刻。式(22)表示场桥的空闲时间。式(23)表示场桥的平均作业量。式(24)表示场桥 λ 的作业量。

为解决目标函数 F_1、F_2、F_3、F_4 和 F_5 量纲不统一的问题，首先生成了 N 个不同的箱位分配与场桥调度协同方案；随后，针对这 N 个方案，计算了以下 5 个加总性能指标作为评估基础，分别为：完成最迟分配任务的场桥作业时间之和 $\sum_{n=1}^{N} \mathcal{Z}_n$，AGV 在堆场处于等待状态的总时间之和 $\sum_{n=1}^{N} \mathcal{W}_n$，箱区总翻箱量之和 $\sum_{n=1}^{N} \mathcal{U}_n$，场桥作业量总方差之和 $\sum_{n=1}^{N} \mathcal{P}_n$，场桥非装卸时间之和 $\sum_{n=1}^{N} \mathcal{X}_n$，无量纲化后的目标函数如式(25)～(29)所示：

$$F'_1 := \frac{F_1}{\sum_{n=1}^{N} \mathcal{Z}_n}, \tag{25}$$

$$F'_2 := \frac{F_2}{\sum_{n=1}^{N} \mathcal{W}_n}, \tag{26}$$

$$F'_3 := \frac{F_3}{\sum_{n=2}^{N} \mathcal{U}_n}, \tag{27}$$

$$F'_4 := \frac{F_4}{\sum_{n=1}^{N} \mathcal{P}_n}, \tag{28}$$

$$F'_5 := \frac{F_5}{\sum_{n=1}^{N} \mathcal{X}_n}。 \tag{29}$$

对多个目标函数进行无量纲化后，为将多目标函数整合为单目标函数，采用线性加权法，根据各目标的重要性分配权重系数，记为 ω_1、ω_2、ω_3、ω_4 和 ω_5，满足 $\omega_1+\omega_2+\omega_3+\omega_4+\omega_5=1$，且 $\omega_1,\omega_2,\omega_3,\omega_4,\omega_5 \geqslant 0$。转化后的单目标函数为：

$$\min F = \omega_1 F'_1 + \omega_2 F'_2 + \omega_3 F'_3 + \omega_4 F'_4 + \omega_5 F'_5。$$

权重系数的具体取值可依据堆场作业的实际研究侧重点进行调整，从而使优化结果更符合实际操作需求。

2　Actor-Critic 算法设计

本文采用基于深度强化学习框架的 Actor-Critic 对上一节所建模型进行求解。Actor-Critic 由 Actor(策略网络)基于状态生成动作概率并执行，Critic(价值网络)评估动作的长期价值，二者通过策略梯度和时序差分协同优化，实现高效决策，在双智能体架构中，可分层协作(先分配箱位再调度场桥)，通过奖励反馈共同提升整体性能。

2.1 状态空间

本文考虑集装箱的箱位分配和场桥调度问题,集装箱和场桥的状态设计如下。

1)集装箱状态。为统一表示,所有任务箱(含出口箱与待提出场箱)均采用1个三元组数据模型进行形式化描述,该模型包含以下3个属性:[当前位置,目标位置,优先级别]。

2)场桥状态。为实现对箱区内多台场桥的全局状态监控,系统使用一个三元组状态向量进行形式化建模,该向量包含以下3个组件:[场桥位置,指派任务,完成时间],其中场桥的位置映射至具体贝位与栈。

2.2 动作空间

1)箱位分配动作。堆场箱位分配问题的核心是:针对每个进场出口箱,将"选择最优堆存位置"的决策,定义为一个决策变量,取值范围构成该问题的动作空间(或可行解空间):

$$\alpha_1=\langle\omega,\gamma,\varepsilon,\theta\rangle\in\otimes\times\beta\times\mathcal{R}_\gamma\times\mathcal{R}_{\gamma\varepsilon},$$

其中:ω 为集装箱唯一标识,且 $\omega\in\otimes$,$\otimes$ 表示集装箱位置坐标(贝、栈、层)的集合;γ 为目标贝位,且 $\gamma\in\beta$;ε 为目标栈号,受限于贝位 γ 的可行列集合,且 $\varepsilon\in\mathcal{R}_\gamma$;$\theta$ 为目标层数,受限于贝位 γ 的栈 ε 的可行层集合,且 $\theta\in\mathcal{R}_{\gamma\varepsilon}$。

2)场桥调度动作。在场桥调度问题中,动作设计为场桥选择待作业任务箱的决策,动作空间可表示为:

$$\alpha_2=\langle\lambda,\omega\rangle\in\kappa\times T_\lambda,$$

其中:λ 为场桥编号,且 $\lambda\in\{1,2\}$;ω 为可分配任务箱集合,且 $\omega\in T_{\lambda p}$。

在动作选择阶段,系统智能体动态评估2台场桥的任务完成时间,据此决策下一台投入运行的场桥。进而,为避免设备间的空间冲突,智能体会对候选任务施加空间隔离约束,即执行区域必须与另一台场桥的作业区保持安全距离,以此滤出可行的后续任务。

3)面向动作空间的智能体探索策略配置。强化学习的决策困境在于"探索"与"利用"的权衡。"利用"旨在基于当前认知做出最优决策,而"探索"则致力于拓宽认知边界。在此背景下,SoftMax算法提供了一种高效的探索方案:它并非盲目尝试,而是根据各动作已获得的平均奖励值作为分配选择权重,并进行概率抽样。这使得智能体既能倾向当前已知的优选项,又不失对长远潜在高回报选项的关注。SoftMax函数在强化学习中的具体应用形式如下所示:

$$P(a\mid s)=\frac{\mathrm{e}^{\frac{Q(s,a)}{\tau}}}{\sum_{a'}\mathrm{e}^{\frac{Q(s,a')}{\tau}}},$$

其中:$P(a\mid s)$ 表示策略中用于计算状态 s 下选择动作 a 的概率,$Q(s,a)$ 表示在状态 s 下采取动作 a 的动作-价值函数,衡量了在状态 s 时选择动作 a 的优劣程度,值越大表示该动作越优。τ 是温度参数,它控制着探索与利用的平衡。τ 的值越大,各动作被选择的概率越趋于平均,智能体更倾向于探索;τ 的值越小,高价值动作被选择的概率越大,智能体更倾向于利用已有经验。分子 $\mathrm{e}^{\frac{Q(s,a)}{\tau}}$ 对动作价值 $Q(s,a)$ 进行指数变换,放大不同动作价值之间的差异程度。分母 $\sum_{a'}\mathrm{e}^{\frac{Q(s,a')}{\tau}}$ 是对所有可能动作 a' 的 $\mathrm{e}^{\frac{Q(s,a')}{\tau}}$ 求和,起到归一化作用,确保计算出的 $P(a\mid s)$ 是一个合法概率,所有动作选择概率之和为1。

2.3 奖惩函数的设计

在深度强化学习中,智能体根据动作效果从环境获得奖励。本模型将箱位分配与场桥调度的合理性量化为一个综合奖励值,该值由5个指标共同决定:

1)最迟完成装船任务的场桥作业时间越短,奖励越大。

2)箱区总翻箱量越小则奖励越高,这是通过评估上下层出口箱的优先级差值来实现的。差值为正时,给予正向奖励;为负时(预示翻箱),则施加惩罚。

3)AGV在堆场等待总时间最小。当AGV等待时间越短,奖励越大。

4)各场桥作业任务量的方差最小,当每个场桥作业量越均衡,奖励越大。

5)场桥的非装卸时间最少,当场桥空闲时间和跑位时间越少,奖励越大。

针对本文目标函数最小化问题,对5个目标设置加权系数进行归一化规范处理,将任意输入映射至[0,1]值域范围内。

总目标为:$F=m_1F_1+m_2F_2+m_3F_3+m_4F_4+m_5F_5$,奖惩函数可设置为:

$$r=\frac{1}{F}=\frac{1}{m_1F_1+m_2F_2+m_3F_3+m_4F_4+m_5F_5}。$$

2.4　神经网络设计

针对箱位分配问题，Actor 网络输入层为当前智能体强化学习的状态值[当前位置，目标位置，优先级别]，该神经网络模型的拓扑结构定义如下：输入层由 3 个神经元节点构成，用以接收状态信息；输出层则生成一个维度与所有可选动作数量相同的向量，该向量经由 SoftMax 函数处理，被转化为一个有效的动作概率分布。此分布作为策略函数，直接决定了目标箱的分配决策。其中，隐藏层的规模设计是一个关键的超参数，它需要在模型表达能力（求解精度）与计算效率（训练时长）之间寻求最佳平衡，数值需通过实验谨慎确定。为避免因隐藏层神经元过多而导致的训练效率低下、过拟合风险增高及陷入局部最优等问题，本研究针对 3 个状态特征的具体情况，选择了 3 层隐藏层与每层 32 个神经元的网络结构。此配置在确保模型具备足够表达能力的同时，也有效控制了计算成本。Actor 深度神经网络结构如图 5 所示。

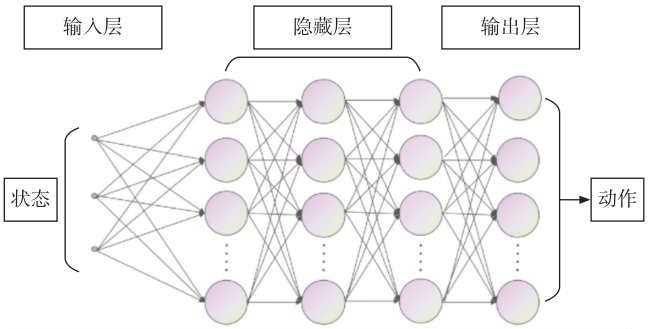


图 5　Actor 深度神经网络结构

Fig. 5　Structure of Actor deep neural network

在协同架构中，Critic 网络作为策略评估器，输入层接收包含状态与即时奖励的四维信号，进而通过单一输出节点计算时序差分（temporal difference，TD）误差。该误差作为核心的评估信号，不仅精准量化了当前策略的优劣，更作为策略梯度反馈至 Actor 网络，指导其参数更新。与之协同的 Actor 网络，输入被定义为由场桥位置、指派任务与完成时间构成的状态矩阵。为确保架构统一，隐藏层设计与 Critic 网络及箱位分配模型保持同构。Critic 深度神经网络结构如图 6 所示。

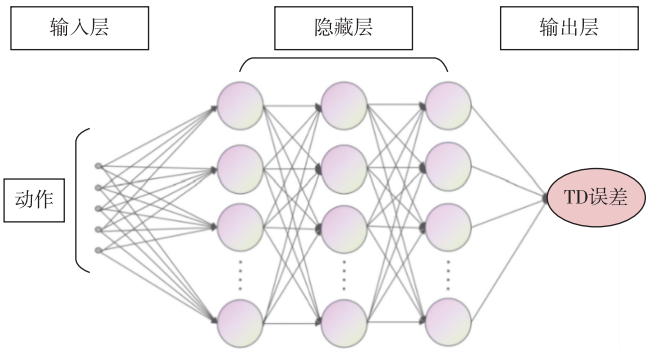


图 6　Critic 深度神经网络结构

Fig. 6　Structure of Critic deep neural network

2.5　Actor-Critic 算法流程

基于 Actor-Critic 协同框架的智能体决策伪代码如下。

输入：最大训练轮数 $T=5\ 000$，最大步数 L（等于任务队列长度），经验回放缓存区容量 $B=10\ 000$，批次大小 $N=32$，Actor 网络学习率 $R_{\mathrm{Al}}=0.001$，Critic 网络学习率 $R_{\mathrm{Cl}}=0.01$，奖励折扣因子 $\xi=0.9$，探索率 $E=0.1$。

输出：训练好的 Actor 网络参数 θ 和 Critic 网络参数 ω，场桥调度方案 S，箱位分配结果 A。

初始化

步骤 1：随机初始化 Actor 网络参数 θ 和 Critic 网络参数 ω。

步骤2:初始化经验回放缓存区 D,容量为 B。

步骤3:初始化目标网络 $\theta' \leftarrow \theta, \omega' \leftarrow \omega$。

主训练循环 for $e=1$ to T do

步骤1:重置环境,获取初始状态 s_1。

步骤2:初始化本轮累积奖励 $r_{\text{total}} \leftarrow 0$。

步骤3:时序交互循环:for $t=1$ to L do

步骤4:动作选择:根据当前策略 π_θ 和状态 s_t 选择动作 a_t。

以概率 ε 随机探索一个合法动作,否则根据 Actor 网络输出的动作概率分布 $\pi_\theta(a \mid s_t)$ 采样动作 a_t。

步骤5:执行动作:在模拟环境中执行动作 a_t(包括箱位分配与场桥任务指派)。

步骤6:观察环境:环境返回下一个状态 s_{t+1}、即时奖励 r_t 和终止标志 done。

步骤7:存储经验:将转移元组(s_t, a_t, r_t, s_{t+1},done)存入经验回放缓存区 D。

步骤8:状态更新:$s_t \leftarrow s_{t+1}$。

步骤9:累积奖励:$r_{\text{total}} \leftarrow r_{\text{total}} + r_t$。

步骤10:记录调度:将动作 a_t 及对应的决策细节(如场桥ID、目标箱位)记录到 S 中。

步骤11:网络参数更新:

若缓存区中样本数足以构成一个批次,则从 D 中随机采样 N 个转移元组。

更新 Critic 网络:对每个样本,计算目标值 $y_i = r_i + \xi * V_{\omega'}(s_{i+1})$,并最小化 Critic 网络的均方误差损失:$L(\omega) = \frac{1}{N} * \Sigma_i (y_i - V_\omega(s_i))^2$。

更新 Actor 网络:使用策略梯度上升更新 Actor 参数:$\theta \leftarrow \theta + \alpha_{\text{Actor}} * \nabla_\theta J(\theta)$,其中:$\nabla_\theta J(\theta) \approx \frac{1}{N} * \Sigma_i [\nabla_\theta \log \pi_\theta(a_i \mid s_i) * A(s_i \mid a_i)]$,优势函数 $A(s_i \mid a_i)$ 可由 Critic 网络输出的 TD 误差 $\delta_i = y_i - V_\omega(s_i)$ 近似。

更新目标网络:$\theta' \leftarrow \tau\theta + (1-\tau)\theta'$,$\omega' \leftarrow \tau\omega + (1-\tau)\omega'$。

输出结果

步骤12:随存最终的网络参数 θ 和 ω。

步骤13:解析最后一次训练的 S,生成最终的场桥调度方案与箱位分配结果 A。

步骤14:返回 θ, ω, S, A。

3 实验验证

3.1 实验基础设计

本文针对第3代现代化内河港口果园港码头堆场的情景设计了不同规模的问题求解。箱区的规模为40个贝位×6个堆垛×4个堆层,每个贝预留2个箱位以满足翻箱需要,即每个贝位包含22个可堆存箱位。设某箱区有50个集装箱装卸任务。相关参数设置见表2。

表2 相关参数

Tab. 2 Related parameters

参数表示	数值	参数表示	数值
箱区规模{$\gamma, \varepsilon, \theta$}	40×6×4	场桥跨列移动耗时/s	2
场桥数量/台	2	场桥提放一个出口箱所需时间/s	30
单位贝位长度/m	7	场桥间的安全距离/m	14
场桥跨贝位运行耗时/s	5		

2台场桥的初始贝位号分别为18和40,50个待堆存出口箱的整体优先级信息如表3所示。各贝位初始堆存量信息如表4所示。鉴于箱区内贝位总数庞大,为清晰呈现,本系统仅截取前10个贝位作为代表,以展示堆存状态。每个贝位的堆存信息采用1个4位的二进制状态编码进行表征,每1位对应1个层位。其中,编码“1”

代表该层存在集装箱，而“0”则标识该层处于空置状态，如表 4 中的第 1 列第 2 行“1100”表示的是在第 1 贝第 2 栈的第 1 层和第 2 层有集装箱，而第 1 贝第 2 栈的第 3 层和第 4 层无集装箱。

表 3　待堆存出口箱信息

Tab. 3　Information of export containers to be stacked

任务箱编号	整体优先级	任务箱编号	整体优先级	任务箱编号	整体优先级	任务箱编号	整体优先级	任务箱编号	整体优先级
1	23	11	13	21	22	31	41	41	43
2	34	12	21	22	21	32	34	42	14
3	11	13	31	23	34	33	21	43	21
4	42	14	12	24	44	34	23	44	34
5	24	15	23	25	11	35	12	45	42
6	13	16	24	26	11	36	32	46	22
7	12	17	13	27	31	37	14	47	14
8	32	18	23	28	24	38	13	48	32
9	24	19	23	29	24	39	21	49	44
10	14	20	11	30	22	40	24	50	12

表 4　箱区部分箱位的初始堆存状态

Tab. 4　Initial stacking status of parital slots in the container yard

栈	贝									
	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	γ_7	γ_8	γ_9	γ_{10}
ε_1	1000	1100	1100	1100	1100	1000	1000	1110	1100	1100
ε_2	1100	1000	1100	1000	1000	1000	1111	1110	1000	1000
ε_3	1110	1100	0000	1000	1110	1000	1000	1100	1110	1100
ε_4	1000	1100	1100	1100	1000	1110	1000	1111	1100	0000
ε_5	0000	1000	1000	1000	1100	1000	1100	1000	1000	1110
ε_6	1000	0000	1100	1100	1000	1000	1000	1000	1100	1100

3.2　实验求解

本文的实验环境为 PyTorch 2.5.1 Pyhton 3.12. GPU：RTX4090(24 GB)。目标函数 F_1、F_2、F_3、F_4 和 F_5 所对应的权重参数值分别设置为 $\omega_1=0.3$，$\omega_2=0.2$，$\omega_3=0.3$，$\omega_4=0.1$，$\omega_5=0.1$。Actor-Critic 算法相关参数设置见表 5。

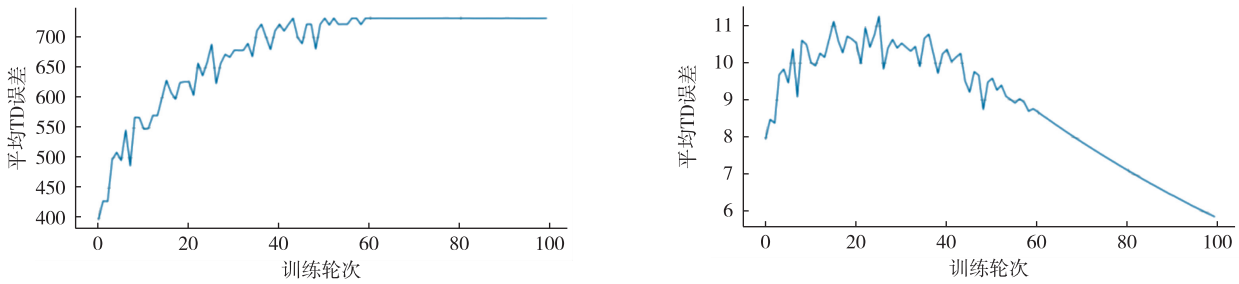
表 5　Actor-Critic 算法所用参数

Tab. 5　Parameters used in the Actor-Critic algorithm

参数	取值	参数	取值
Actor 网络学习因子 R_{Al}	0.001	神经网络节点个数	32
Critic 网络学习因子 R_{Cl}	0.01	迭代次数	5 000
TD 误差时奖励折扣因子 ξ	0.9		

在 Actor-Critic 算法框架下，本设计采用正向奖励函数：奖励值为正，且数值大小与系统性能正相关。具体而言，更高的奖励值意味着更优的调度方案。即场桥作业耗时更短、AGV 等待时长更少、翻箱次数更低、场桥非装卸时间更少、场桥作业量更均衡，进而能体现出场桥调度和箱区分配的协同优化效果更好。根据这些性能指标，绘制的奖惩值收敛曲线和时间差分误差收敛曲线如图 7 所示。奖惩值收敛曲线表明：在训练初期，即训练轮次为 0 到 1 000 左右，总奖励呈快速上升趋势。说明智能体在前期通过与环境交互，快速学习到有效的策略，获

得的奖励大幅增加。之后奖励值虽有波动,但整体围绕一个较高的值波动并趋于平稳,大约在730。体现了Actor-Critic算法在训练过程中,智能体从对环境不熟悉到逐渐掌握较优策略的过程。稳定的奖励值意味着算法找到相对稳定且较优的决策方式来解决堆场调度优化问题。时间差分误差收敛曲线表明:在训练开始阶段,即训练轮次为0到1 000左右,平均TD误差有较大波动且数值较高,这是因为算法刚开始学习,对状态—价值的估计还不准确。随着训练轮次增加,误差呈持续下降趋势,在后期,训练轮次为4 000~5 000,下降速度虽变缓但仍在降低,说明算法对状态价值的估计越来越准确。TD误差作为衡量Critic网络对状态价值估计准确性的指标,误差不断减小,表明Critic网络能更精准地评估Actor网络所采取动作的优劣,为Actor网络提供更有价值的反馈,进而辅助Actor网络更好地调整策略,提升算法整体性能。



a 奖惩值收敛曲线　　b TD误差收敛曲线

注:图中的训练轮次的次数是每50次迭代记录1次奖惩值。

图7 奖惩值收敛曲线和TD误差收敛曲线

Fig. 7 Reward convergence curve and temporal difference error convergence curve

基于Actor-Critic算法的训练生成了箱位分配(表6、图8)与场桥调度(表7、图9)方案。箱位分配结果确认了50个出口箱均被合理分配,且无一违反堆叠约束。在图8中,不同颜色表示不同层,紫色代表第1层,蓝色代表第2层,绿色代表第3层,黄色代表第4层。根据不同颜色,构成箱位分配实验结果,所有箱体均被分配至独立且空闲的箱位,确保了堆存的有序性。同时,箱体在箱区内实现了均衡分布,这为2台场桥的负载均衡奠定了基础,从而有效缩短了场桥大车的移动距离,并降低了设备间的作业干扰。场桥调度方案结果如表7和图9所示,表7表明每台场桥的作业量为25个集装箱,满足作业均衡的目标。场桥作业轨迹图图9表明,2台场桥的路径无交点,任意时刻2台场桥所在的贝位保持一定的间距,说明求解的箱位分配与场桥调度方案没有出现场桥间的相互影响或干扰,满足场桥间安全距离的要求。

表6 箱位分配结果表

Tab. 6 Container slot allocation result table

任务箱编号	箱位分配方案	任务箱编号	箱位分配方案	任务箱编号	箱位分配方案	任务箱编号	箱位分配方案	任务箱编号	箱位分配方案
1	15-4-2	11	11-2-1	21	16-4-3	31	39-1-4	41	30-3-1
2	17-6-3	12	18-6-4	22	25-6-2	32	24-2-3	42	12-3-1
3	15-6-4	13	19-5-4	23	27-6-2	33	16-5-1	43	28-3-4
4	35-6-2	14	6-6-2	24	35-3-4	34	30-5-2	44	26-6-1
5	24-2-2	15	26-2-2	25	27-4-2	35	3-6-4	45	40-4-1
6	6-6-3	16	20-2-3	26	8-1-4	36	21-2-2	46	29-5-2
7	12-1-2	17	12-1-1	27	20-2-2	37	15-4-3	47	3-6-3
8	24-5-2	18	19-4-3	28	30-4-3	38	4-3-2	48	28-2-2
9	16-6-3	19	25-1-4	29	26-2-3	39	18-3-3	49	39-4-4
10	14-3-3	20	1-3-4	30	15-5-3	40	23-4-4	50	8-2-4

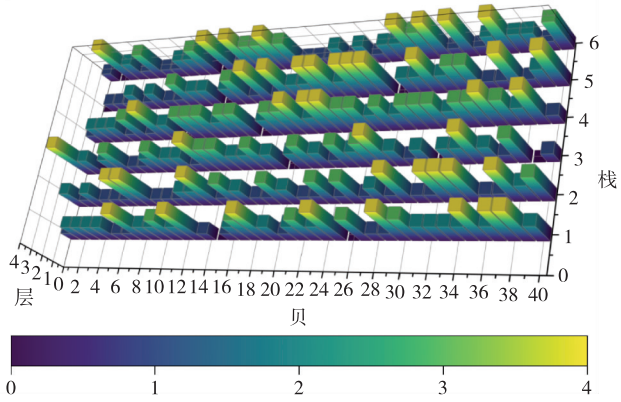


图 8　箱位分配实验结果图

Fig. 8　Container slot allocation experimental result diagram

表 7　场桥调度方案

Tab. 7　Yard bridge scheduling plan

场桥作业顺序	场桥 1		场桥 2		场桥作业顺序	场桥 1		场桥 2	
	出口箱编号	贝位号	出口箱编号	贝位号		出口箱编号	贝位号	出口箱编号	贝位号
1	12	18	24	35	14	11	11	22	25
2	39	18	28	30	15	50	8	19	25
3	18	19	48	28	16	6	6	8	24
4	2	17	25	27	17	35	3	32	24
5	21	16	23	27	18	20	1	36	21
6	33	16	4	35	19	47	3	27	20
7	30	15	34	30	20	38	4	5	24
8	37	15	41	30	21	42	12	15	26
9	1	15	43	28	22	14	6	40	23
10	9	16	31	39	23	26	8	44	26
11	7	12	46	29	24	13	19	49	39
12	10	14	16	20	25	3	15	45	40
13	17	12	29	26					

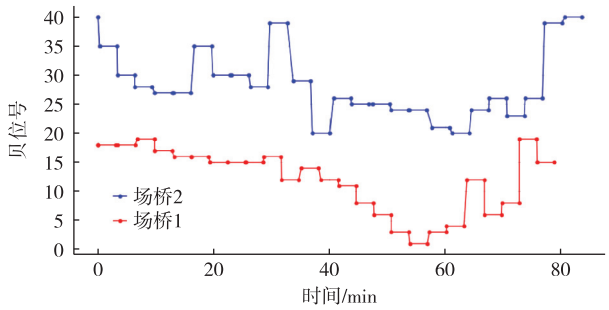


图 9　场桥作业轨迹图

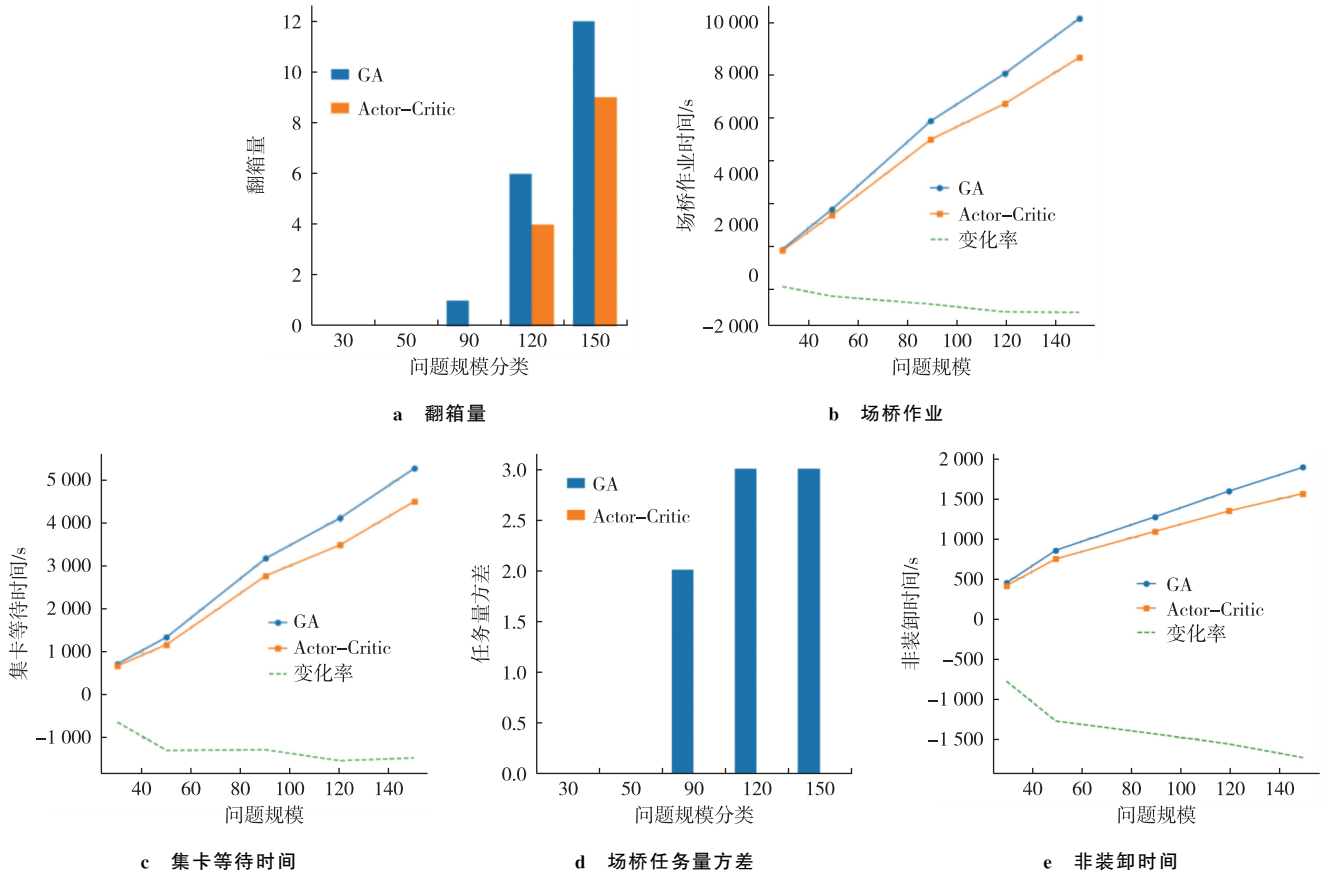
Fig. 9　Yard bride operation trajectory diagram

4　算例结果对比分析

4.1　不同规模下的 GA 和 Actor-Critic 算法结果对比分析

为了与传统 GA 对比，本文分别设计了规模为 30、50、90、120 和 150 的算例。设定由 2 台场桥进行作业，箱

区在初始时刻已有集装箱堆存。分别使用 Actor-Critic 算法和 GA 进行求解,所得结果如图 10 和表 8 所示。在翻箱量上,在较小规模(30 和 50)中,Actor-Critic 和 GA 都表现良好,但随着问题规模增大,GA 的重定位次数呈指数增长,而 Actor-Critic 增长平缓,说明 Actor-Critic 策略更稳定。在场桥作业时间上,二者作业时间随规模增大而呈线性增长,但 Actor-Critic 的斜率更低,说明该策略能更高效地分配任务。相较 GA 缩短 10%～20%,且规模越大优化越明显。在集卡等待时间上,等待时间与任务规模正相关,但 Actor-Critic 的优化百分比保持稳定,说明它的策略能有效降低拥堵。且对比 GA 缩短作业时间 15%～20%,优化效果稳定。在场桥任务量方差上,Actor-Critic 方案在任务量方差控制上表现出色,始终为 0,稳定性强。在非装卸时间上,二者优化百分比随规模增大而提高,但 Actor-Critic 算法减少无效时间 15%～20%,直接提升作业效率。在求解速度上,由于 GA 依赖种群迭代,计算成本随规模增大而呈指数增长。Actor-Critic 则通过策略网络直接输出动作,避免重复评估,因此 Actor-Critic 求解速度比 GA 更快更稳定。



a 翻箱量 b 场桥作业 c 集卡等待时间 d 场桥任务量方差 e 非装卸时间

图 10 不同算法对比图

Fig. 10 Comparison chart of different algorithms

通过以上对比表明 Actor-Critic 算法在降低作业成本、提升效率、均衡任务分配方面全面优于 GA,尤其在大规模问题中优势明显。

4.2 与已有文献核心指标对比分析

为了进一步验证本文箱位分配与场桥调度协同模型的实用性和有效性,本文与已有文献进行对比分析。在具有相同任务数的情况下,对比不同模型和算法得出的各类指标。

文献[9]以均衡各场桥作业任务量和减少场桥的非装卸时间为目标,建立混合整数规划模型,并设计 GA 求解箱位分配和场桥调度问题。文献[11]提出基于 Actor-Critic 的堆场策划与场桥调度方案用以减少场桥作业时间、内集卡等待时间和箱区翻箱量。相较文献[9-11],本文同时考虑场桥作业时间方差、非装卸时间、场桥作业时间、内集卡等待时间和翻箱量因素作为目标函数,建立了新的混合整数规划模型,并采用基于 Actor-Critic 的算法进行求解,与文献[9-11]的对比结果见表 9。

表 8　不同算法对比表

Tab. 8　Comparison table of different algorithms

实验序号	任务规模	GA						Actor-Critic					
		翻箱量	场桥作业/s	集卡等待/s	任务量方差	非装卸时间/s	CPU 耗时/s	翻箱量	场桥作业/s	集卡等待/s	任务量方差	非装卸时间/s	CPU 耗时/s
1	30	0	962	726	0	432	36.2	0	913	681	0	398	25.4
2	50	0	2 541.5	1 352	0	840	58.7	0	2 316	1 179	0	732	42.1
3	90	1	6 035	3 182	2	1 256.4	132.6	0	5 309.2	2 779	0	1 074.8	88.3
4	120	6	7 924	4 119	3	1 578.6	174.3	4	6 728	3 492.8	0	1 330.4	109.2
5	150	12	10 083	5 272	3	1 876.2	235.1	9	8 545	4 505	0	1 549.3	146.5

表 9　不同模型算法对比表

Tab. 9　Comparison table of different model algorithms

目标函数	文献[9]	文献[11]	本文
任务量方差	0		0
非装卸作业时间/min	17.52		12.2
任务完成总时间/min	167.52		123.1
平均作业效率/(箱·h^{-1})	17.9		23.98
翻箱量		1	0
场桥作业时间/s		2 560	2 316
集卡等待时间/s		1 400	1 179
CPU 耗时/s		50.2	42.1

由表 9 可知，进行对比的任务量方差均为 0，说明在任务分配的均衡性上，二者表现相同，都能实现任务的均匀分配，避免出现任务量过度集中在某些执行单元的情况。但在减少非装卸时间、缩短总任务时长及提升作业效率方面均优于文献[9]，说明本文减少了无效等待或移动时间，整体作业效率更高。由表 9 可知，与文献[11]相比，本文所得结果：场桥作业时间减少 244 s、集卡等待时间减少 221 s、重定位次数减少 1 次、CPU 耗时缩短 8.1 s。说明在基于本文所建的新混合整数规划模型下，整体作业效率和资源利用上表现更优。

5　结论

本文在场桥分区域作业模式下，研究了集装箱码头箱位分配与场桥调度的协同优化问题。建立了融合场桥作业均衡、场桥非装卸时间、场桥作业时间、内集卡等待时间和翻箱量在内的混合整数规划模型。相比已有模型，该模型从更多角度衡量场桥调度和箱位分配的协同关系，更能反映场桥调度和箱位分配的协同优化效果。针对所建模型，本文采用了基于 Actor-Critic 框架的深度强化学习方法进行求解。最后通过与已有算法和已有箱位分配与场桥调度结果实验对比，仿真实验表明，相较于传统 GA，基于 Actor-Critic 的算法在小规模问题上能够快速准确地求解，同时在大规模问题求解上相较于 GA 展现了明显的优越性。与已有文献协同优化结果相比，一方面本文箱位分配结果精确到具体箱位，也就是具体到贝、栈、层；另一方面，本文所得结果在保持场桥作业均衡下，能有效降低场桥作业的完成时间和非装卸时间、缩短 AGV 的等待时长和翻箱量。

本研究提出的协同优化框架与深度强化学习求解方法可推广至传统非自动化集装箱码头。对于传统非自动化集装箱码头，仅需将本文中的 AGV 替换为普通的集卡车，调度逻辑不变。此外，通过调整输入状态维度，可以适配具有不同箱区布置的港口码头。但需要指出的是，本研究构建的混合整数规划模型及求解算法是基于确定性环境假设的，即 AGV 到达时间、场桥作业效率等参数均为已知且固定。这一假设构成了当前模型的主要局限：在实际作业中，若发生 AGV 延迟、场桥故障或临时插单等动态不确定性，原定优化方案的性能可能下降。因此，未来研究将引入随机优化等方法以增强调度策略的抗干扰能力与系统鲁棒性。

参考文献:

[1] 刘瑞麟. 确定任务视角下集装箱堆场箱位分配研究[D]. 大连:大连海事大学,2023.
Liu Ruilin. Study on space allocation of container yard from the perspective of definite task[D]. Dalian: Dalian Maritime University, 2023.

[2] 余明珠,梁卓斌,金波. 基于仿真优化的自动化集装箱码头出口箱箱位分配[J]. 运筹与管理,2023,32(5):181-189.
Yu Mingzhu, Liang Zhuobin, Jin Bo. Storage space allocation for outbound containers in automated container terminals using simulation-based optimization[J]. Operations Research and Management Science, 2023, 32(5): 181-189.

[3] Yang X R, Hu H T, Cheng C. Flexible yard space allocation plan for new type of automated container terminal equipped with unilateral-cantilever rail-mounted gantry cranes[J]. Advanced Engineering Informatics, 2023, 58: 102193.

[4] 郑云峰,刘瑞麟,饶本顺,等. 确定任务视角下集装箱堆场箱位分配研究[J]. 重庆交通大学学报(自然科学版),2023,42(7):120-127.
Zheng Yunfeng, Liu Ruilin, Rao Benshun, et al. Location allocation of container yard from the perspective of determining tasks[J]. Journal of Chongqing Jiaotong University (Natural Sciences), 2023, 42(7): 120-127.

[5] He J L, Tan C M, Zhang Y T. Yard crane scheduling problem in a container terminal considering risk caused by uncertainty [J]. Advanced Engineering Informatics, 2019, 39: 14-24.

[6] Quintero-Araújo C L, Caballero-Villalobos J P, Montoya-Torres J R. A multi-objective genetic algorithm for yard crane scheduling in container terminals[J]. Computers & Industrial Engineering, 2019, 129: 404-416.

[7] Chang D F, Jiang Z H, Yan W, et al. Developing a dynamic rolling-horizon decision strategy for yard crane scheduling[J]. Advanced Engineering Informatics, 2011, 25(3): 485-494.

[8] 郑红星,王杰,姚琳. 预倒出口箱区的多场桥调度优化[J]. 运筹与管理,2020,29(12):13-22.
Zheng Hongxing, Wang Jie, Yao Lin. Optimization of multiple crane yards scheduling in pre-rehandling export block[J]. Operations Research and Management Science, 2020, 29(12): 13-22.

[9] 范厚明,姚茜,马梦知. 多场桥分区域平衡策划下的集装箱堆场箱位分配问题[J]. 控制与决策,2016,31(9):1603-1608.
Fan Houming, Yao Qian, Ma Mengzhi. Storage space allocation based on regional workload balance planning of multiple yard cranes in container terminal yard[J]. Control and Decision, 2016, 31(9): 1603-1608.

[10] 马晓斌. 送箱集卡失约下多箱区箱位分配及场桥调度协同优化[D]. 大连:大连海事大学,2024.
Ma Xiaobin. Container allocation in multi-blocks and collaborative optimization of yard crane dispatching in non-engagement of container delivery truck[D]. Dalian: Dalian Maritime University, 2024.

[11] 董良才,岳涵词,王维娟. 基于Actor-Critic深度强化学习的堆场分布式资源联合计划与调度优化研究[J/OL]. 计算机工程,2025-01-03. https://doi.org/10.19678/j.issn.1000-3428.00EC0070379.
Dong Liangcai, Yue Hanci, Wang Weijuan. Research on the joint planning and scheduling optimization of distributed resources in the yard based on actor-critic deep reinforcement learning[J/OL]. Computer Engineering, 2025-01-03. https://doi.org/10.19678/j.issn.1000-3428.00EC0070379.

[12] Chen J, Huang S, Zhou X. Collaborative optimization of container allocation and yard crane scheduling with time windows[J]. Ocean & Coastal Management, 2023, 232: 106532.

[13] 王小寒,靳志宏. 海铁联运港站共享堆场的箱区分配及转场联合优化[J]. 交通运输系统工程与信息,2024,24(5):283-294.
Wang Xiaohan, Jin Zhihong. Joint optimization of block allocation and transfer for shared yards at sea-rail intermodal container terminals[J]. Journal of Transportation Systems Engineering and Information Technology, 2024, 24(5): 283-294.

[14] 王无印,黄子钊,庄子龙,等. 基于深度强化学习的自动化码头堆场场桥调度方法[J]. 机械工程学报,2024,60(6):44-57.
Wang Wuyin, Huang Zizhao, Zhuang Zilong, et al. Yard crane scheduling method based on deep reinforcement learning for the automated container terminal[J]. Journal of Mechanical Engineering, 2024, 60(6): 44-57.

[15] Liu X, Li Z, Yang Y. A deep reinforcement learning approach for coordinated yard crane scheduling and container storage allocation[J]. Advanced Engineering Informatics, 2022, 54: 101724.

[16] Kim K H, Hong G P. A heuristic rule for relocating blocks[J]. Computers & Operations Research, 2006, 33(4): 940-954.

Special Section on Modern Sorting Theory and Job Scheduling Optimization

Coordinated Optimization of Container Yard Slot Allocation and Quay Crane Scheduling Based on Actor-Critic

LIU Zhanlong[1], WANG Yi[2], SUN Xiaochi[3], LI Yifeng[1], ZHANG Xingong[1]

(1. Chongqing National Center for Applied Mathematics, Chongqing Normal University, Chongqing 401331;
2. Chongqing Port Co., Ltd., Chongqing 400025;
3. Chongqing Guoyuan Container Terminal Co., Ltd., Chongqing 401133, China)

Abstract: As a key node in international trade, the operational efficiency of container terminals directly affects the turnaround time of ships and logistics costs. Bay allocation and yard crane scheduling are two crucial aspects in container terminal operations, which are interdependent and influence each other. Therefore, this paper proposes a method based on the deep reinforcement learning Actor-Critic framework to study the collaborative optimization of bay allocation and yard crane scheduling, aiming to achieve collaborative decision-making in these two areas. On the one hand, a mixed-integer programming model is established considering factors such as yard crane operation balance, non-operational time of yard cranes, operation time of yard cranes, waiting time of internal container trucks, and container reshuffling volume. On the other hand, by constructing the state space of the container yard, the action space of yard cranes, and a multi-objective reward function model, a solution algorithm based on the deep reinforcement learning Actor-Critic framework is proposed. Through comparative analysis of different-scale examples, it is verified that the proposed algorithm outperforms the traditional genetic algorithm in terms of solution quality. Compared with existing research results, under the condition of maintaining the operation balance of yard cranes, the new algorithm can effectively reduce the operation and non-operational time of yard cranes, shorten the waiting time of automated guided vehicles, and lower the container reshuffling rate.

Keywords: slot allocation; yard crane scheduling; deep reinforcement learning; Actor-Critic; mixed-integer programming model; container terminal

(责任编辑 黄 颖)