E-Commerce Letters 电子商务评论, 2025, 14(12), 5022-5032
Published Online December 2025 in Hans. https://www.hanspub.org/journal/ecl
https://doi.org/10.12677/ecl.2025.14124456

Hans汉斯

基于深度强化学习的云印刷服务平台动态定价与智能派单机制研究

金懿心[1]，乔俊伟[2]，党亚峥[3*]

[1]上海理工大学出版学院，上海
[2]上海出版印刷高等专科学校印刷包装工程系，上海
[3]上海理工大学管理学院，上海

收稿日期：2025年11月4日；录用日期：2025年11月17日；发布日期：2025年12月29日

摘　要

随着按需经济的深化与制造业的数字化转型，云印刷服务平台面临着需求高度波动、产能资源异构和决策实时性要求高的三重挑战。现有运营策略多依赖静态规则，在理论上缺乏应对复杂动态环境的系统性决策框架，难以实现全局最优化。为填补此理论空白，本文立足于运营管理与人工智能交叉领域，提出一种基于深度强化学习的动态定价与智能派单联合优化框架。首先，从系统工程理论出发，将云印刷平台的运营过程形式化地建模为马尔可夫决策过程(Markov Decision Process, MDP)，系统性地刻画了市场状态、订单流与工厂负载间的复杂动态交互关系。其次，为求解该模型，本文设计了一套双层决策机制：上层采用基于深度确定性策略梯度(DDPG)的动态定价模型，将定价问题视为连续动作空间下的最优控制问题，以实现平台长期收益最大化；下层则引入多智能体协作范式，构建了基于多智能体近端策略优化(MAPPO)的智能派单模型，将各印刷厂视为独立决策的智能体，通过集中式训练与去中心化执行实现资源的高效协同调度。最后，通过高保真仿真平台进行的大量实验表明，相较于基线方法，本框架在订单接受率、平台总收益和平均完工时间等关键绩效指标上均取得显著提升。研究不仅为云印刷领域的智能化运营提供了有效的解决方案，更在理论层面验证了将强化学习与多智能体系统理论相结合，用以解决复杂生产服务系统联合优化问题的可行性与优越性，为相关领域的研究提供了新的理论视角与实践参考。

关键词

云印刷，动态定价，智能派单，深度强化学习，多智能体系统，资源优化

Research on Dynamic Pricing and Intelligent Dispatching Mechanism for Cloud Printing Service Platforms Based on Deep Reinforcement Learning

Yixin Jin[1], Junwei Qiao[2], Yazheng Dang[3*]

*通讯作者。

文章引用: 金懿心, 乔俊伟, 党亚峥. 基于深度强化学习的云印刷服务平台动态定价与智能派单机制研究[J]. 电子商务评论, 2025, 14(12): 5022-5032. DOI: 10.12677/ecl.2025.14124456

[1]College of Publishing, University of Shanghai for Science and Technology, Shanghai
[2]Department of Printing and Packaging Engineering, Shanghai Publishing and Printing College, Shanghai
[3]Business School, University of Shanghai for Science and Technology, Shanghai

Received: November 4, 2025; accepted: November 17, 2025; published: December 29, 2025

Abstract

Driven by the deepening of the on-demand economy and the digital transformation of manufacturing, cloud printing service platforms face threefold challenges: high demand volatility, heterogeneous production resources, and the imperative for real-time decision-making. Prevailing operational strategies, which predominantly rely on static rules, theoretically lack a systematic framework to address such complex dynamic environments, thus failing to achieve global optimization. To fill this theoretical gap, this paper, positioned at the intersection of operations management and artificial intelligence, proposes a joint optimization framework for dynamic pricing and intelligent dispatching based on deep reinforcement learning. First, from a system engineering perspective, the operational process of the cloud printing platform is formally modeled as a Markov Decision Process (MDP), which systematically characterizes the complex dynamic interactions among market status, order flow, and factory load. Second, to solve this model, a two-tiered decision mechanism is designed: the upper tier employs a dynamic pricing model based on Deep Deterministic Policy Gradient (DDPG), treating the pricing problem as an optimal control problem in a continuous action space to maximize the platform's long-term revenue. The lower tier introduces a multi-agent collaborative paradigm by constructing an intelligent dispatching model based on Multi-Agent Proximal Policy Optimization (MAPPO), which regards each printing factory as an independent agent to achieve efficient collaborative scheduling through centralized training and decentralized execution. Finally, extensive experiments on a high-fidelity simulation platform demonstrate that, compared to baseline methods, the proposed framework achieves significant improvements in key performance indicators such as order acceptance rate, total platform revenue, and average completion time. This research not only provides an effective solution for the intelligent operation of the cloud printing industry but also theoretically validates the feasibility and superiority of integrating reinforcement learning with multi-agent systems theory for solving joint optimization problems in complex production-service systems, offering a novel theoretical perspective and practical reference for future research in related fields.

Keywords

Cloud Printing, Dynamic Pricing, Intelligent Dispatching, Deep Reinforcement Learning, Multi-Agent System, Resource Optimization

1. 引言

在工业 4.0 与服务型制造的浪潮推动下，传统印刷业正经历着深刻的数字化转型。云印刷作为一种整合分散产能、满足个性化按需生产的新型网络化服务模式，其市场规模正以前所未有的速度扩张。它打破了传统生产的时空限制，将复杂的印刷供应链转化为一个动态的、数据驱动的在线服务平台。然而，

在这种高效灵活的表象之下，平台运营决策的复杂性也急剧增加。订单流的随机性、生产资源的异构性与服务质量要求的多样性共同构成了一个高维、动态、不确定的决策环境，对平台的运营效率与收益管理(Operations Efficiency and Revenue Management)构成了严峻挑战。当前，云印刷平台的运营决策在理论与实践层面均存在显著瓶颈。实践中，平台普遍采用基于成本加成的静态定价策略和基于就近或低价原则的规则派单系统。这类策略虽然简单直观，但在理论上属于开环控制(Open-loop Control)，无法根据实时的市场供需、工厂负载等动态信息进行闭环反馈与自适应调整，导致资源错配与利润流失。学术界虽有应用运筹学模型进行定价或调度的研究，但这些模型往往依赖于对市场需求的强先验假设，难以刻画真实环境的随机性与复杂性。近年来，深度强化学习(DRL)作为解决复杂序列决策问题的强大范式，在网约车调度、能源管理等领域展现了巨大潜力，但其在云印刷这一生产与服务紧密耦合场景的应用尚属空白。特别是，如何构建一个能够对定价和派单两个核心杠杆进行联合优化的、系统性的智能决策理论框架，已成为该领域亟待突破的关键科学问题。深度强化学习(Deep Reinforcement Learning, DRL)作为一种在复杂序列决策问题上表现卓越的技术[1]，通过智能体与环境的交互试错来学习最优策略，已在网约车动态定价与派单[2]、智能车队管理[3]等领域展现出巨大潜力。这些场景在资源供需动态匹配与智能决策方面与云印刷平台具有高度的相似性，为解决云印刷领域的运营难题提供了新的技术路径。受此启发，本研究旨在将深度强化学习理论引入云印刷服务领域，以应对其核心的动态定价与智能派单挑战。本文构建了一个动态定价与智能派单的联合优化框架，将定价问题建模为连续动作空间下的强化学习任务，并提出一种基于深度确定性策略梯度(DDPG)的动态定价模型以实现价格的精细化、实时化调整。同时，为高效调度异构的工厂资源，本文将复杂的派单问题分解为多智能体协同决策问题，并设计了一种基于多智能体近端策略优化(MAPPO)的智能派单算法。最后，通过搭建仿真环境对所提机制的有效性进行验证，以期为云印刷平台的智能化升级和效率提升提供新的理论方法与实践参考。

2. 相关工作

本研究的核心在于应用深度强化学习技术解决云印刷平台中的动态定价与智能派单联合优化问题。因此，本章节将从动态定价、智能派单与调度两个核心研究领域展开综述，并分析深度学习技术在云印刷及相关运营决策场景中的应用现状，从而明确本文的研究空白与贡献。

2.1. 动态定价研究

动态定价作为收益管理的关键组成部分，旨在根据动态的市场环境实时调整价格以最大化收益。传统方法多依赖于对市场需求函数做出较强假设的经济学模型与运筹优化理论，在处理高度不确定的复杂环境时面临挑战，尤其是在需要考虑平台竞争与需求随机性的场景下[4]。近年来，数据驱动的方法，特别是深度强化学习(DRL)，为此领域带来了新的突破[5]。DRL 无需精确的环境模型，能够通过与环境的直接交互来学习复杂的市场动态，从而制定出更具适应性的定价策略。然而，现有研究多集中于零售或出行服务，鲜有针对云印刷这类与生产能力和资源负载紧密耦合的场景进行定制化研究。

2.2. 现有决策模型的理论局限性

尽管现有研究已取得一定进展，但其在应对云印刷这类复杂生产服务系统时，仍存在显著的理 论局限性。首先，基于传统运筹学的优化模型，如混合整数规划(MIP)，虽然能提供最优解的理论保证，但其模型构建往往依赖于对市场需求、生产时间等参数的确定性或简单随机分布假设。这与云印刷场景中高度动态和不确定的现实环境相悖，导致模型鲁棒性差，难以适应实时决策需求。其次，早期的启发式[6]

或元启发式算法[7]，虽提升了求解效率，但其本质上是基于规则的搜索策略，缺乏从数据中学习和自适应演化的能力，难以发现反直觉的最优策略，其解的质量也缺乏理论保证。再次，现有将机器学习应用于该领域的研究多集中于单点优化，例如，仅预测订单量或优化单一的派单环节。这种“割裂式”的优化忽略了定价与派单两个核心决策之间强烈的耦合关系——价格直接影响订单接受率，进而改变派单系统的负载分布。缺乏一个联合优化的系统性理论框架，是导致当前智能化水平无法实现全局最优的根本原因。因此，本研究的核心动机，正是要突破这种单点优化的局限，从一个系统性的、闭环的视角，构建一个数据驱动的联合决策理论框架。

2.3. 智能派单与调度研究

智能派单本质上是一个大规模的组合优化与资源分配问题。传统方法如启发式算法在面对高动态、大规模的调度需求时，往往难以保证决策的实时性和最优性。深度强化学习为此类序列决策问题提供了强大的解决范式[8]。对于云印刷平台中涉及多个异构工厂协同的场景，多智能体强化学习(MARL)展现出独特的优势[9]。MARL 允许将每个决策单元(如印刷工厂)建模为独立的智能体，通过协同学习来完成复杂的全局调度任务，已在数据中心资源管理、车联网资源分配[10] [11]等分布式场景中得到成功应用。国内学者也对值分布的多智能体方法等前沿方向进行了探索[12]。尽管如此，将 MARL 应用于印刷任务调度，并考虑印刷工艺、设备负载、交付时效等多重约束的研究 尚不充分，尤其是在需要实时在线决策的复杂场景中[13]。

2.4. 云印刷领域的智能化研究现状

目前，人工智能技术在云印刷领域的应用尚处于初级阶段，研究主要集中在生产流程的自动化，如利用计算机视觉进行印前文件分析与印后质量检测等。在更为核心的平台运营决策层面，尤其是动态定价与智能派单这两个紧密耦合的问题上，相关研究十分匮乏。现有平台运营模式仍较为传统，其智能化决策的潜力远未被发掘。因此，本研究旨在填补这一空白，首次尝试将深度强化学习与多智能体系统理论相结合，为云印刷服务平台的联合定价与派单问题提供一个系统性的、数据驱动的解决方案，推动行业的智能化运营升级。

3. 模型与方法

本章旨在对云印刷平台中的动态定价与智能派单联合优化问题进行系统性的建模，并详细阐述所提出的基于深度强化学习的决策模型。首先，将介绍系统的总体框架，从宏观层面展示数据流与决策模块间的交互关系。其次，为将该序列决策问题转化为强化学习可解的形式，将采用马尔可夫决策过程(MDP)进行形式化定义，明确状态、动作、奖励等核心要素。最后，将深入介绍分别用于解决动态定价和智能派单两个子问题的核心算法模型——基于 DDPG 的定价模型与基于 MAPPO 的多智能体派单模型。为系统性地解决云印刷平台中的动态定价与智能派单联合优化问题，本文设计了一个数据驱动的闭环智能决策框架，如图 1 所示。该框架遵循感知－决策－执行－反馈(Perception-Decision-Execution-Feedback, PDEF)的控制论思想，将平台运营视为一个持续学习与优化的自适应系统。该框架始于感知与状态表征层，负责实时采集并整合来自市场、订单和工厂的多源异构数据，构建系统在 t 时刻的全面状态向量 S_t。这些信息随即输入框架的核心，即联合优化决策层，该层包含动态定价与智能派单两个相互耦合的决策模块，协同生成联合动作 A_t。随后，在物理执行与环境交互层，印刷工厂与用户分别根据决策执行生产与响应报价，其结果构成了环境反馈。最终，学习与更新反馈回路利用环境反馈(Reward, R_t)和新的状态 S_t+1 持续训练并迭代决策模型，从而形成一个不断自我演进和优化的闭环系统。

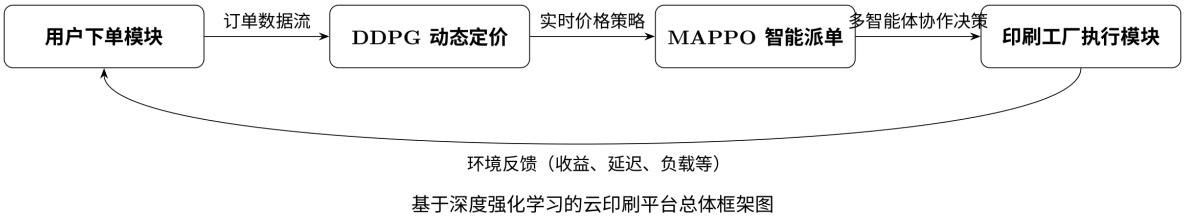

Figure 1. Overall framework of the deep reinforcement learning-based cloud printing platform
图 1. 基于深度强化学习的云印刷平台总体框架图

3.1. 问题形式化定义

我们将云印刷平台的运营过程建模为一个马尔可夫决策过程(MDP)，其核心要素定义如下：

- 状态(State, S)：状态 $s_t \in S$ 是对 t 时刻系统环境的向量化描述，由市场状态(S^M)、订单状态(S^O)和工厂状态(S^F)拼接而成。其中，S^M 包含时间编码、近期订单平均到达率等；S^O 包含待处理订单数量、平均复杂度、平均紧急度等；S^F 为各工厂的归一化队列长度、设备平均负载等组成的向量。
- 动作(Action, A)：联合动作 $a_t \in A$ 由定价和派单两部分组成。定价动作 $A_t^P \in [-1,1]$ 是一个连续值，代表价格调整因子。派单动作为多智能体联合动作 $A_t^D \in \left[A_{t,1}^D, \cdots, A_{t,N_f}^D\right]$，其中每个智能体(工厂)的动作 $A_{t,i}^D$ 为对待分配订单的偏好排序或选择。
- 奖励(Reward, R)：为综合评估策略的长期效益，奖励函数 r_t 设计为一个多目标的加权和：

$$r_t = \alpha r_{\text{revenue}} - \beta c_{\text{wait}} - \gamma c_{\text{overload}} \tag{1}$$

其中，r_{revenue} 为平台收益，c_{wait} 和 c_{overload} 分别为订单平均等待时间和工厂超载的惩罚项。权重系数 α，β，γ 通过超参数搜索确定，该设计使得模型在收益最大化的同时兼顾用户体验与系统平衡性。该奖励函数的设计，本质上是一个多目标优化问题在强化学习框架下的体现，其背后蕴含了平台运营的关键理论权衡。平台收益(R_{revenue})是核心的经济目标，直接关联收益管理理论(Revenue Management)。然而，单纯追求短期收益最大化可能导致接受大量高价低效订单，损害长期用户体验。因此，我们引入订单平均等待时间(C_{wait})作为惩罚项，这源于排队论(Queuing Theory)和服务质量(QoS)管理，旨在保障用户满意度和平台声誉。同时，工厂超载(C_{overload})惩罚项则根植于生产与运作管理(Production and Operations Management)的理论，旨在维持生产系统的稳定性，避免因个别工厂过载导致的系统性崩溃和交付延迟。权重系数 α，β，γ 的设定，为平台在“效率(收益)”“公平(等待)”与“稳定(负载)”这三个核心运营目标之间进行战略取舍提供了量化调节的手段。

状态转移概率(Transition Probability, P)：$P\left(s_{t+1} \middle| s_t, a_t\right)$ 由真实环境的内在随机性决定，在无模型(model-free)的强化学习框架中无需显式定义。

3.2. 云印刷运营系统建模与数据抽象

在将云印刷平台的运营过程形式化为 MDP 之前，我们首先对其核心业务实体与数据流进行系统性建模。该系统主要包含四个核心实体：用户(User)、订单(Order)、工厂(Factory)和平台(Platform)。平台的运营状态本质上是这四大实体在时间序列上的动态演化。为了构建 MDP 的状态空间，我们从原始的业务数据中进行数据抽象。市场状态(S^M)通过聚合近期订单到达率、平均订单金额、用户拒绝率等时序数据来表征；订单状态(S^O)则由待处理订单队列的长度、订单的平均复杂度、平均紧急度等特征构成；而工厂状态(S^F)则通过各工厂的实时队列长度、设备平均负载、特殊工艺能力占用率等向量化表示。这些经过抽象和归一化的多维数据，共同构成了在任意时刻 t 精确描述整个系统环境的状态向量 S_t，这是实现智能决策的数据基石。

3.3. 模型假设与局限性分析

尽管本文构建的 MDP 模型力求还原真实的云印刷场景，但为了保证计算的可行性并聚焦于核心的定价与派单决策，我们在建模过程中引入了以下假设与简化，并在此分析其对实际应用可能产生的影响：

用户行为理性化假设：在用户行为建模中，我们基于 Logit 模型假设用户对价格和交付时间具有理性的敏感度曲线，且一旦下单不进行撤单操作。现实中用户可能受品牌忠诚度或非理性因素影响，且存在中途撤单风险。这可能导致模型在极端情况下高估订单转化率，但在长周期统计规律上，该假设仍能有效反映市场供需变化的总体趋势。

物流配送抽象化：鉴于本文聚焦于生产与定价环节，我们将物流配送简化为固定的时间窗口和成本函数，未考虑路况突发拥堵或配送员调度细节。这一简化可能导致在地理位置跨度极大的订单处理上，对总完工时间的预估存在偏差。未来可通过接入实时物流 API 数据来增强模型对“生产－配送”全链路时效的感知能力。

设备状态稳定性假设：假设在决策周期内，印刷设备不发生突发性灾难故障，仅考虑正常的排队与加工时间。在实际生产中，设备故障可能导致产能突降。虽然 MAPPO 机制能通过感知队列积压自动减少向故障工厂派单，但缺乏专门的故障应对机制可能导致短期内的响应滞后。

3.4. 基于 DDPG 的动态定价模型

定价决策本质上是在一个连续区间内寻找最优价格调整因子，以最大化长期累计收益，这可以被视为一个典型的最优控制问题(Optimal Control Problem)。深度确定性策略梯度(DDPG)算法[14]作为一种专门解决连续动作空间强化学习问题的 Actor-Critic 方法，与本问题的性质高度契合。DDPG 基于 Actor-Critic 框架，其中 Actor 网络负责输出确定的价格动作，Critic 网络负责评估该动作的长期价值(Q 值)，并通过 Critic 的指导来优化 Actor 的策略，其核心概念结构如图 2 所示。

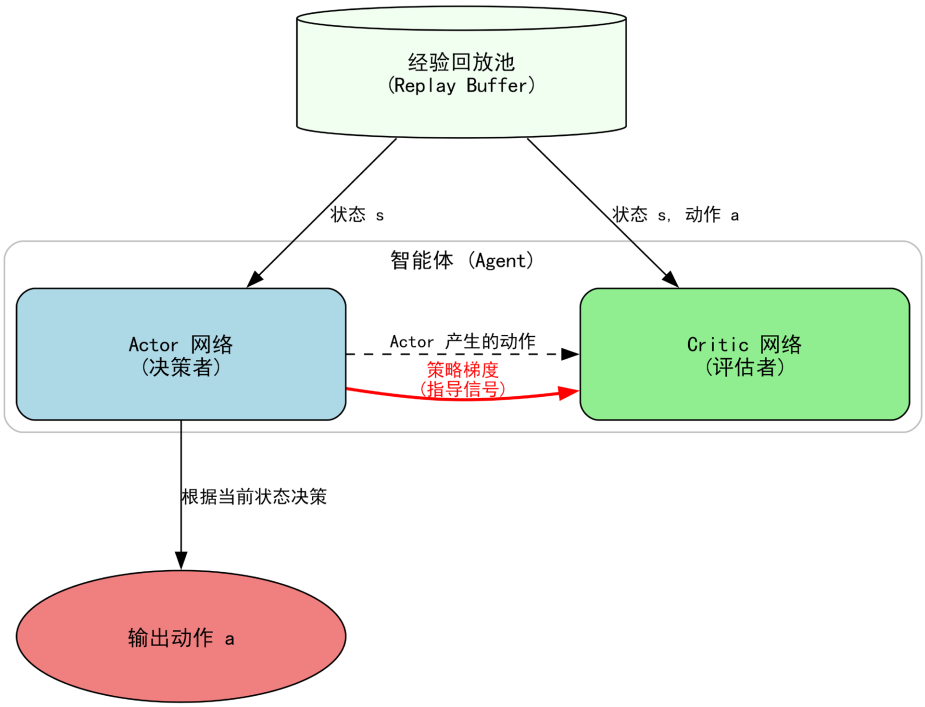

Figure 2. Core conceptual architecture of the DDPG model
图 2. DDPG 模型核心概念结构

3.5. 基于 MAPPO 的智能派单模型

云印刷平台的派单任务涉及为多个订单在多个异构工厂间进行最优匹配，是一个大规模组合优化问题。若采用中心化控制器，将面临随工厂和订单数量指数增长的“维度灾难”。为此，本文引入分布式人工智能与多智能体系统(Multi-Agent Systems, MAS)理论，将每个工厂建模为一个独立的、以实现全局目标为导向的智能体(Agent)。在众多多智能体强化学习(MARL)算法中，本文选用在协同任务中表现出色的多智能体近端策略优化(MAPPO)算法[15] [16]进行训练。其核心思想为“集中式训练，去中心化执行”(CTDE)，从而兼顾了全局最优与执行效率，模型示意图如图 3 所示。

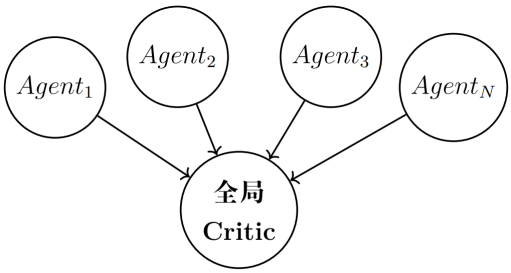

Figure 3. Schematic of the MAPPO-based multi-agent dispatching model
图 3. 基于 MAPPO 的多智能体派单模型示意图

4. 实验设计与结果分析

4.1. 仿真环境与实验设置

为验证本文所提方法的有效性，我们基于 Python 及 SimPy 离散事件仿真库搭建了一个高保真的云印刷平台仿真环境。

• 订单生成：模拟订单流遵循非齐次泊松过程，以模拟一天内订单到达率的高峰与低谷周期。每个订单的属性(如复杂度、印刷量、地理位置)从预设分布中随机采样。

• 用户行为：用户是否接受报价基于 Logit 模型进行模拟，即价格越高，用户接受并支付的概率越低。

• 工厂建模：平台连接 $N_f = 10$ 个异构工厂，其异构性体现在处理速度、产能上限及特殊工艺能力上。

实验对比算法包括：FP-RB (固定定价 + 规则派单)、DP-RB (DDPG 定价 + 规则派单)、FP-IB (固定定价 + MAPPO 派单)以及本文提出的 DDPG-MAPPO 联合模型。主要实验参数设置如表 1 所示。

Table 1. Key experimental parameter settings
表 1. 主要实验参数设置

参数名称	取值
状态维度	48
折扣因子 γ	0.99
学习率(Actor/Critic)	1e−4/1e−3
批量大小(Batch Size)	256
经验池容量	1,000,000
目标网络软更新系数 τ	0.005
PPO 裁前系数 ϵ	0.2
训练迭代次数	500,000

4.2. 性能对比与分析

各算法在关键性能指标上的对比结果如表 2 所示。从表中可以看出，本文提出的 DDPGMAPPO 联合优化模型在所有指标上均表现最佳。具体分析：1) 对比 DP-RB 与 FP-RB，引入动态定价能显著提升平台总收益和订单接受率，证明了 DDPG 在捕捉市场动态方面的有效性。2) 对比 FP-IB 与 FP-RB，引入智能派单能大幅降低订单平均完成时间并改善负载均衡，证明了 MAPPO 在异构资源调度上的优越性。3) DDPG-MAPPO 模型结合了二者优势，通过协同优化实现了全局最优，其总收益相较次优的 DP-RB 模型提升了 14.1%，平均完成时间则与最优的 FP-IB 模型持平，展现了“1 + 1 > 2”的耦合效应。平台累计收益变化曲线(图 4)也直观地验证了本文方法在长期收益上的领先地位。

Table 2. Performance comparison of different algorithms across various metrics
表 2. 不同算法在各项指标上的性能对比

算法	总收益(万元)	接受率(%)	平均完成时间(h)	负载均衡度(方差)
FP-RB	120.5	65.8	24.5	15.8
DP-RB	145.2	72.3	23.9	14.5
FP-IB	138.6	66.1	**18.2**	8.2
DDPG-MAPPO	**165.7**	**75.1**	18.5	**7.5**

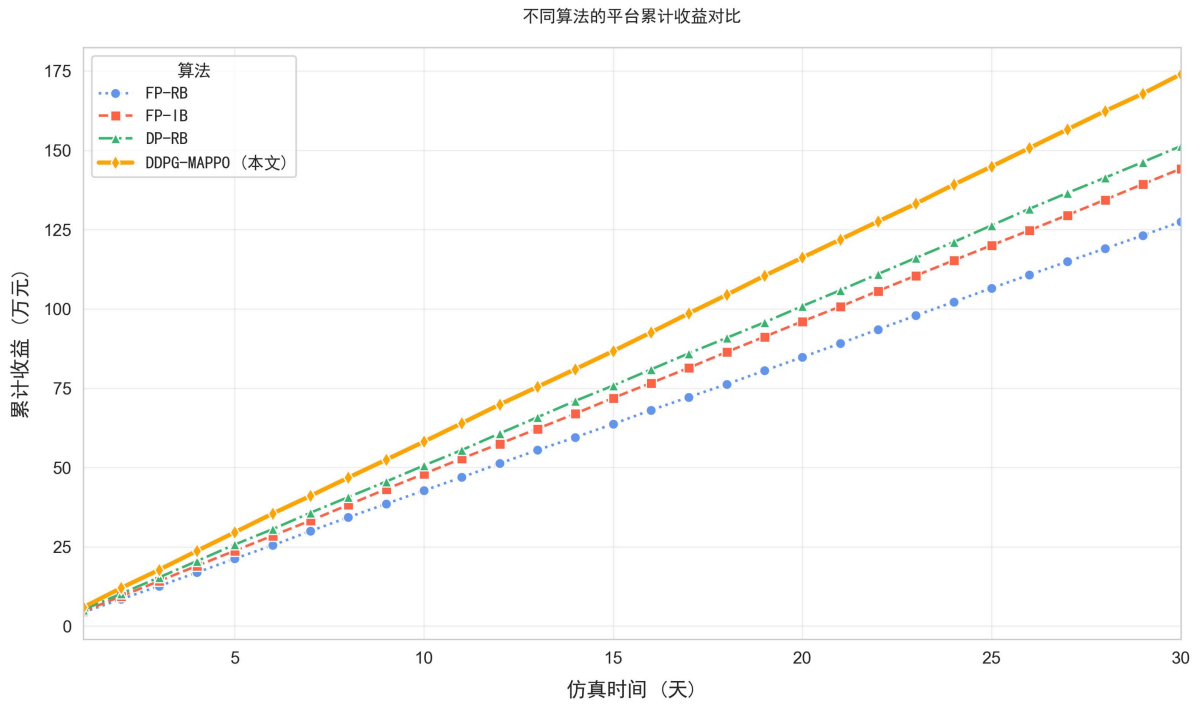

Figure 4. Curve of platform cumulative revenue over time
图 4. 平台累计收益随时间变化曲线

4.3. 典型场景下的微观决策分析

为了深入探究 DDPG-MAPPO 框架“智能”的内在逻辑，我们从仿真数据中截取了两个典型的运营片段(订单高峰期与工厂负载失衡期)，可视化展示本文模型与基线策略(FP-RB)的微观决策差异。

4.3.1. 应对订单高峰期的动态定价策略

图 5 展示了仿真时间 13:00~15:00 订单爆发期间的系统响应。如图 5(a)所示，当订单到达率激增(蓝色虚线)时，基线策略(FP-RB)维持固定价格，导致订单照单全收，工厂队列迅速积压(红色实线)，平均等待时间飙升。相比之下，本文的 DDPG 模块敏锐捕捉到市场过热信号，迅速将定价因子 A_t^p 上调约 15%~20% (绿色实线)。这一操作虽然短期内轻微降低了接单率，但有效筛选了高价值订单，抑制了低价值冗余需求，使得工厂负载维持在健康水位(黄色实线)，避免了系统拥堵导致的整体服务质量崩塌。

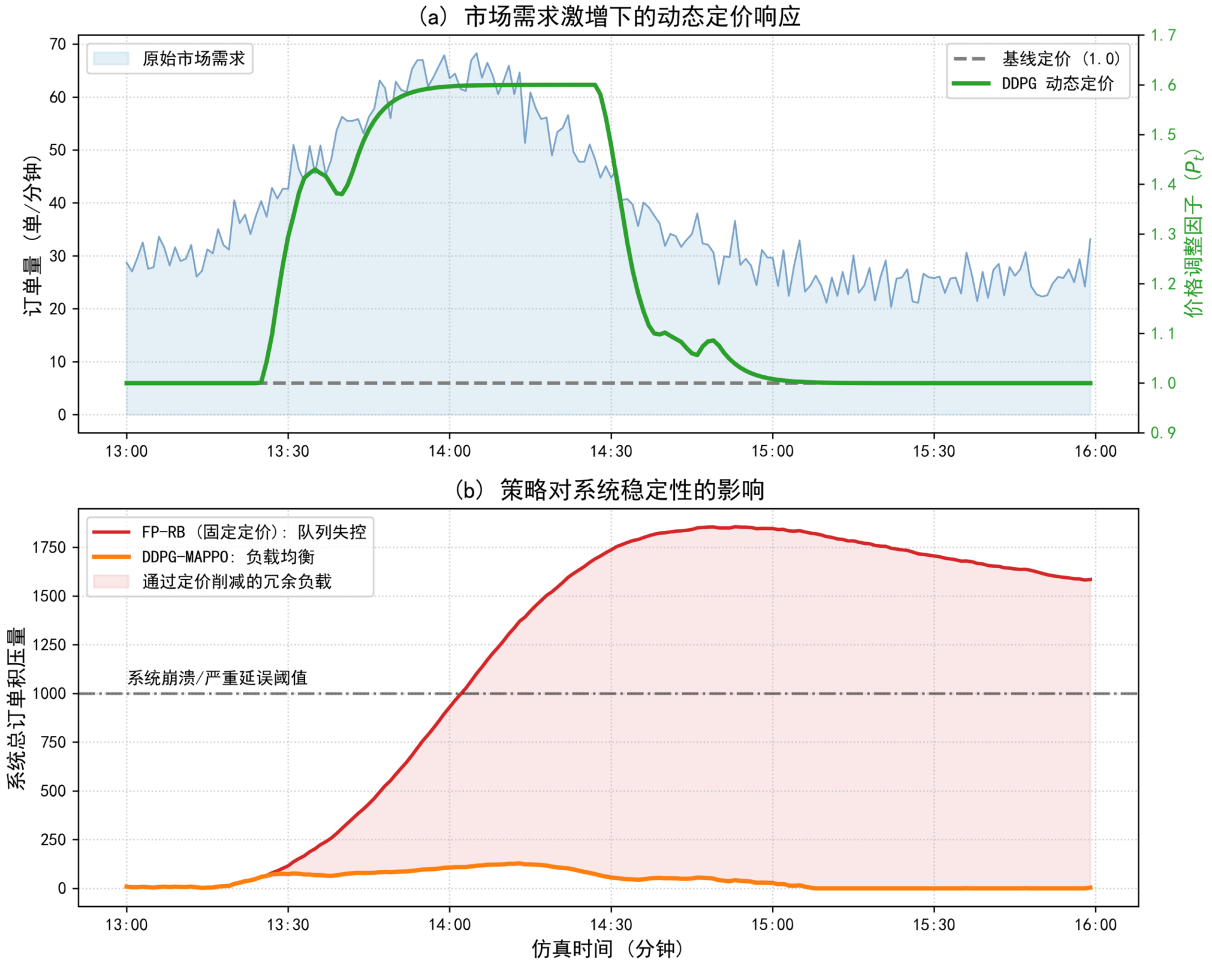

Figure 5. Analysis of dynamic pricing and queue response during order peak periods
图 5. 订单高峰期动态定价与队列响应分析

4.3.2. 应对异构资源瓶颈的智能派单

图 6 展示了某特定工厂(工厂 A，擅长普通印刷但产能较小)负载接近饱和时的派单行为。在传统规则派单(如最近距离优先)下，由于工厂 A 地理位置优越，大量订单仍被持续派往该厂，导致其队列长度远超警戒线，而稍远的工厂 B (产能大)却处于半闲置状态。而在 MAPPO 多智能体系统下，如图 6(b)所示，代表工厂 A 的智能体主动降低了对新订单的竞标权重，而系统通过全局价值函数评估后，将原本属于工厂 A 辐射区的部分大批量订单，智能调度至工厂 B。尽管这增加了少量的物流成本，但显著缩短了 30%以上的排队等待时间，实现了“以空间换时间”的全局最优调度。

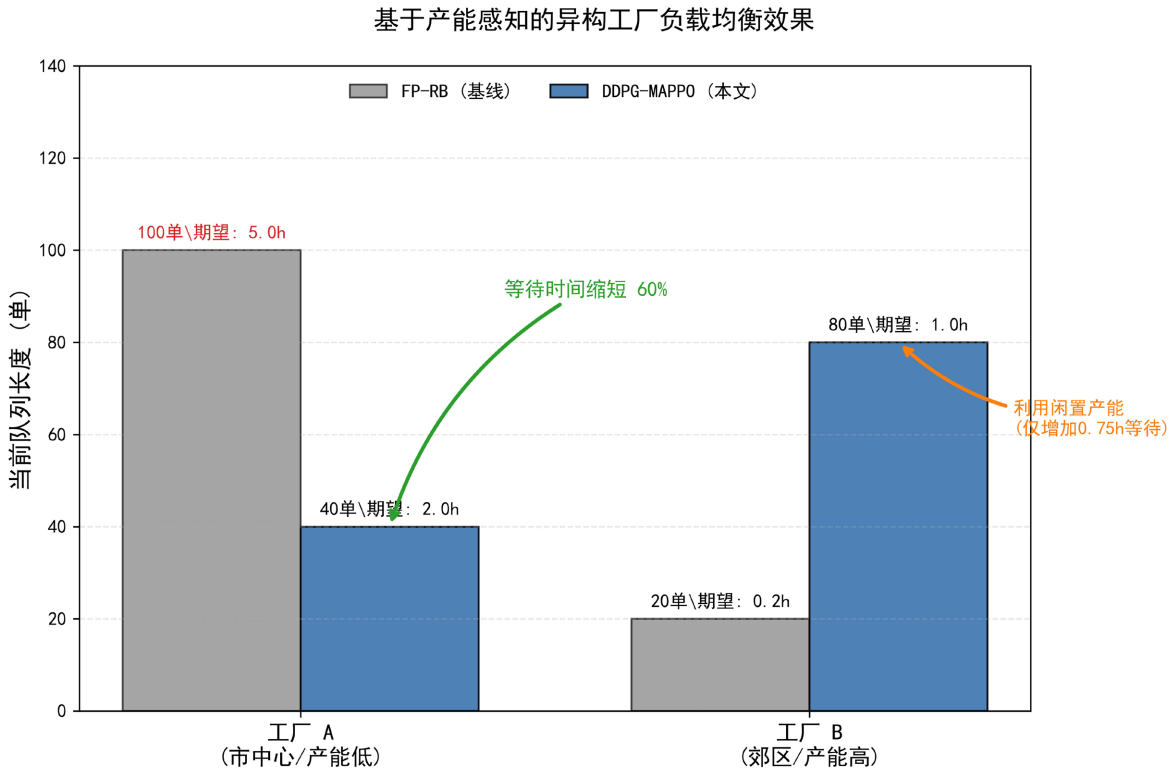

Figure 6. Comparison of intelligent dispatching in factory load imbalance scenarios
图 6. 工厂负载失衡场景下的智能派单对比

4.4. 消融实验分析

为量化模型中各关键组件的贡献，我们设计了消融实验，结果如表 3 所示。实验设置包括：1) 完整模型(DDPG-MAPPO)；2) 移除动态定价(w/o DP)，即采用固定价格；3) 移除智能派单(w/o ID)，即采用基于规则的派单。结果表明，移除任一组件都会导致性能显著下降，其中移除动态定价对总收益影响最大(下降 16.4%)，而移除智能派单对平均完成时间影响最大(恶化 29.2%)。这进一步证明了本文提出的联合优化框架中，动态定价与智能派单两个模块相辅相成，缺一不可。

Table 3. Comparison of ablation study results
表 3. 消融实验结果对比

模型变体	总收益(万元)	平均完成时间(h)
DDPG-MAPPO (完整模型)	**165.7**	**18.5**
w/o Dynamic Pricing (即 FP-IB)	138.6 (−16.4%)	18.2 (−1.6%)
w/o Intelligent Dispatching (即 DP-RB)	145.2 (−12.4%)	23.9 (+29.2%)

5. 结论与展望

本文针对云印刷服务平台运营决策的核心挑战，超越了传统的单点优化思路，创新性地提出并验证了一个动态定价与智能派单的联合优化理论框架。本研究的核心理论贡献在于：成功地将复杂的生产服务系统运营问题，通过马尔可夫决策过程进行了系统性建模；并开创性地融合了基于最优控制思想的连

续决策模型(DDPG)与基于多智能体协作理论的分布式决策模型(MAPPO)，为解决此类耦合决策问题提供了一种全新的、行之有效的研究范式。仿真实验结果充分证明了该框架的有效性和鲁棒性，展现了其在提升平台运营效率与经济效益方面的巨大潜力。这项工作不仅为云印刷行业的智能化升级提供了关键技术支撑，其所构建的“系统建模–联合决策–协同优化”的分析框架，对其他面临类似动态资源分配与定价挑战的领域(如共享制造、云计算服务、智慧物流等)也具有重要的理论借鉴意义和迁移价值。未来的研究工作可从以下几个方面展开：1) 将模型扩展到包含物流配送在内的更完整的供应链条；2) 探索混合动作空间下的端到端联合优化模型；3) 在真实的云印刷平台上进行 A/B 测试，验证其在生产环境中的实际效果，并根据反馈持续迭代优化。

参考文献

[1] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., et al. (2015) Human-Level Control through Deep Reinforcement Learning. *Nature*, **518**, 529-533. https://doi.org/10.1038/nature14236

[2] Wang, Y., Sun, H., Lv, Y., Chang, X. and Wu, J. (2024) Reinforcement Learning-Based Order-Dispatching Optimization in the Ride-Sourcing Service. *Computers & Industrial Engineering*, **192**, Article 110221. https://doi.org/10.1016/j.cie.2024.110221

[3] Hazrathosseini, A. and Moradi Afrapoli, A. (2024) Transition to Intelligent Fleet Management Systems in Open Pit Mines: A Critical Review on Application of Reinforcement-Learning-Based Systems. *Mining Technology*: *Transactions of the Institutions of Mining and Metallurgy*, **133**, 50-73. https://doi.org/10.1177/25726668231222998

[4] 孙中苗, 徐琪. 随机需求下考虑不同竞争情形的网约车平台动态定价[J]. 中国管理科学, 2021, 29(1): 138-148.

[5] Zhu, C., Cheng, C. and Meng, S. (2024) DRL Pricepro: A Deep Reinforcement Learning Framework for Personalized Dynamic Pricing in E-Commerce Platforms with Supply Constraints. *Spectrum of Research*, **4**.

[6] 李煜, 马良. 新型元启发式布谷鸟搜索算法[J]. 系统工程, 2012, 30(8): 64-69.

[7] 赵畅, 刘允刚, 陈琳, 等. 面向元启发式算法的多无人机路径规划现状与展望[J]. 控制与决策, 2022, 37(5): 1102-1115.

[8] 郑小红, 龙军, 蔡志平. 关于网约车订单分配策略的综述[J]. 计算机工程与科学, 2020, 42(7): 1267-1275.

[9] Hady, M.A., Hu, S., Pratama, M., Cao, Z. and Kowalczyk, R. (2025) Multi-Agent Reinforcement Learning for Resources Allocation Optimization: A Survey. *Artificial Intelligence Review*, **58**, Article No. 354. https://doi.org/10.1007/s10462-025-11340-5

[10] Wang, J., He, H., Cha, J., Jeong, I. and Ahn, C. (2025) Multi-Agent Reinforcement Learning for Efficient Resource Allocation in Internet of Vehicles. *Electronics*, **14**, Article 192. https://doi.org/10.3390/electronics14010192

[11] 王辛果, 王昶. 一种采用联邦深度强化学习的车联网资源分配方法[J]. 电讯技术, 2024, 64(7): 1065-1071.

[12] 韩明志, 李宁, 王超. 基于值分布的多智能体强化学习方法[J]. 计算机科学与应用, 2024, 14(4): 201-212.

[13] 武小平, 兰梦月. 网约车实时订单在线选择策略设计与竞争分析[J]. 运筹与管理, 2024, 33(4): 153-158.

[14] Lillicrap, T.P., Hunt, J.J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D. and Wierstra, D. (2015) Continuous Control with Deep Reinforcement Learning. arXiv:1509.02971.

[15] Schulman, J., Wolski, F., Dhariwal, P., Radford, A. and Klimov, O. (2017) Proximal Policy Optimization Algorithms. arXiv:1707.06347.

[16] Yu, C., Velu, A., Vinitsky, E., Gao, J., Wang, Y., Bayen, A. and Wu, Y. (2022) The Surprising Effectiveness of PPO in Cooperative Multi-Agent Games. *Advances in Neural Information Processing Systems*, **35**, 24611-24624.