

# 面向电子商务系统的多业务确定性调度： 一种图神经网络增强的深度强化学习方法

冯佳俊

贵州大学大数据与信息工程学院, 贵州 贵阳

收稿日期: 2025年12月20日; 录用日期: 2025年12月30日; 发布日期: 2026年1月30日

## 摘要

在现代电子商务系统中, 业务请求通常呈现出显著的异构时间特性。不同类型的业务(如在线支付、订单确认、实时推荐与风险控制)具有差异化的服务级别协议(SLA)约束, 其截止时间与时延敏感性存在显著差异。这一特性使得传统基于统一策略的请求调度方法难以同时兼顾系统效率与服务质量: 一方面, 统一调度容易导致高实时性业务请求错过时延约束; 另一方面, 过于保守的资源分配策略又会造成计算与通信资源的低效利用, 增加系统运营成本。针对上述问题, 本文提出了一种异构图神经网络增强的深度强化学习调度框架, 面向云-边-服务节点协同的电子商务业务调度场景。该框架在构建的异构知识图谱中显式建模业务类型、业务请求与计算资源之间的关联关系, 使图神经网络能够有效捕获不同业务请求之间的依赖关系、资源状态动态以及业务类型约束。在此基础上, 引入一种熵正则化自适应调度策略, 在满足关键业务时延与SLA约束的前提下, 实现系统调度性能与资源利用率的稳定优化。

## 关键词

电子商务, 调度, 深度强化学习, 熵正则化

# Multi-Service Deterministic Scheduling for E-Commerce Systems: A Graph Neural Network-Enhanced Deep Reinforcement Learning Approach

Jiajun Feng

College of Big Data and Information Engineering, Guizhou University, Guiyang Guizhou

## Abstract

In modern e-commerce systems, business requests typically exhibit significant heterogeneous temporal characteristics. Different types of business operations (such as online payments, order confirmation, real-time recommendations, and risk control) have differentiated service level agreement (SLA) constraints, with significant differences in deadlines and latency sensitivity. This characteristic makes it difficult for traditional request scheduling methods based on a unified strategy to simultaneously balance system efficiency and service quality: on the one hand, unified scheduling easily leads to high-real-time business requests missing latency constraints; on the other hand, overly conservative resource allocation strategies result in inefficient utilization of computing and communication resources, increasing system operating costs. To address these issues, this paper proposes a heterogeneous graph neural network-enhanced deep reinforcement learning scheduling framework for cloud-edge-service node collaborative e-commerce business scheduling scenarios. This framework explicitly models the relationships between business types, business requests, and computing resources in a constructed heterogeneous knowledge graph, enabling the graph neural network to effectively capture the dependencies between different business requests, dynamic resource states, and business type constraints. Based on this, an entropy-regularized adaptive scheduling strategy is introduced to achieve stable optimization of system scheduling performance and resource utilization while satisfying critical business latency and SLA constraints.

## Keywords

E-Commerce, Scheduling, Deep Reinforcement Learning, Entropy Regularization

---

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

第六代(6G)智能网络的演进显著提升了网络的连接规模与计算分布程度,为大规模、低时延的电子商务应用提供了重要支撑。然而,这一演进也使电子商务系统中的业务调度在复杂性与动态性方面进一步加剧。特别是在高并发与多业务并行的电商场景中,业务请求普遍呈现出显著的异构时间特性:不同类型的业务在时延敏感性与截止时间要求上存在明显差异。若采用统一的调度策略,异构业务的差异性往往被忽略,进而导致整体调度效率下降。尤其是在端-边-云协同的电子商务系统中,该问题更加突出:研究表明,超过80%的业务对毫秒级时延具有严格要求,而在异构业务被同质化处理的情况下,系统容易出现业务截止时间违约或计算与通信资源利用不足等现象。随着6G网络持续推动计算架构向高度分布式与低时延方向演进,如何在满足关键业务服务级别协议(SLA)约束的同时,实现高效、稳定的多业务调度,已成为制约新一代电子商务系统性能与服务质量提升的关键挑战。

本文将具有异构时间约束的多业务请求确定性调度问题建模为一个业务类型感知的异构时空匹配问题。该建模框架能够在统一结构中捕捉业务类型、时间约束、计算能力之间的复杂依赖关系,为异构业务请求的协调调度提供理论基础。同时引入熵正则化策略通过自适应调整探索程度和价值估计,提升调

度策略在动态与对抗环境下的稳定性与收敛效率。

## 2. 研究现状

### 2.1. 基于深度强化学习的业务调度

为了实现业务的确定性传输，有效地在设备之间进行业务调度至关重要。深度强化学习(DRL)在多业务、多设备调度中提供了强有力的支持。

Kang 等人提出了一个基于自适应深度强化学习(ADRL)的业务调度框架，旨在减少业务的平均响应时间和提高资源的 CPU 利用率[1]。与此同时，Wang 等人提出了一种基于 DRL 算法的表示模型，以应对节点和业务动态变化所引发的维度灾难，从而降低能耗[2]。尽管这些 DRL 方法在各种业务调度场景中表现出色，但其依赖于精确的状态表示来学习有效的策略。传统状态表示方法在高维和复杂状态空间中往往难以全面捕捉系统的全局特性，从而限制了策略学习的效率和效果。

为了解决状态表示中的问题，研究者们尝试了多种改进方法。例如，Jiang 等人应用具有无监督学习的相关正则化堆叠自动编码器(2r-SAE)对高维信道质量信息(CQI)数据进行数据压缩和表示，以减少 DRL 的状态空间从而提高学习效率，但这种方法可能会导致信息丢失[3]。在[4]中，Orr 等人提出了一种新的状态空间模型 - 分配排序(AO)，其通过强制执行特定的调度顺序，在状态空间搜索中提供了优势。不过 AO 模型在低通信成本的调度问题上表现不如 ELS 模型，部分原因在于 ELS 模型的剪枝技术和 f 值计算更为有效[5]。提出了一种分层建模方法，通过将非状态空间模型与状态空间模型集成，有效克服了状态爆炸问题，但也带来了额外的计算开销。

尽管这些方法在简化状态空间和提高学习效率方面取得了一定进展，但它们仍存在局限性。在此背景下，研究者们开始探索更为先进的表示学习工具，以进一步提升状态表示的精度和鲁棒性。图神经网络(GNN)作为一种有效的表示学习方法，能够通过建模节点之间的复杂关系，增强对状态空间的建模能力[6]。Liang Liu 等人利用析取图来编码 DRL 状态，并使用提出的 GNN 模型来学习节点嵌入[7]。Peng 等人表明，使用线性处理器的边隐藏表示可以实现与使用非线性处理器的顶点隐藏表示相同的性能，但训练时间和推理时间显著较短[8]。然而，上述方法多基于同质图假设，难以有效刻画现实系统中的异构结构特征。

### 2.2. 电子商务系统服务优化与资源调度研究

电子商务系统服务优化是国内学者长期关注的重要研究方向。相关研究主要围绕业务响应时延、服务质量保障、资源协同与系统稳定性等问题展开。

例如，覃雪莲等人从供应链与平台运作视角出发，构建了面向电子商务场景的服务优化与协同调度模型，以提升系统在高并发条件下的整体服务能力和运行效率[9]。张鹏等人基于大数据分析方法，对电商业务负载特征与服务性能进行建模，提出面向动态需求变化的资源调配与服务优化策略，为电商系统的服务质量保障提供了重要理论支撑[10]。然而，这类研究多依赖规则驱动或静态优化模型，在面对业务类型高度异构、负载快速波动的在线调度场景时，自适应能力仍然有限。

综上所述，现有 DRL 调度研究主要面向通用计算或通信场景，缺乏对电商系统中高并发、业务类型异构及时延约束差异显著等特征的系统性建模与研究；而现有电商服务优化虽在资源协同与时延压缩方面成果丰富，但较少将异质图结构建模与深度强化学习相结合，用于在线调度决策问题。针对上述研究空白，本文提出了一种基于异质图神经网络(HGNN)的深度强化学习调度框架，通过对电商系统中不同类型实体及其交互关系进行统一建模，构建高鲁棒性的系统状态表示，并据此驱动 DRL 智能体学习满足多

业务时延约束的协同调度策略。

### 3. 系统模型

#### 3.1. 网络架构

我们设计了一个多业务多阶段调度框架。该架构采用端-边-云协同范式，通过高速光纤链路互连。端节点称为业务生成设备，负责生成计算业务。边缘服务器称为业务处理设备，负责执行相应的子业务。中心云负责协调整个系统的业务调度。当业务生成设备生成业务时，其会主动向中心云汇报业务的相关信息。同时，业务处理设备也会向中心云汇报它的处理能力及所能处理的业务类型。

每个业务由一组相互依赖的子业务组成，每个子业务映射到具有不同计算功能的设备。并遵循严格的执行顺序。设  $\mathcal{U} = \{1, \dots, u, \dots, U\}$  为业务处理设备的集合， $\mathcal{I} = \{1, \dots, i, \dots, I\}$  表示业务集合， $\mathcal{J} = \{1, \dots, j, \dots, J\}$  表示子业务集合， $\mathcal{F} = \{1, \dots, f, \dots, F\}$  表示子业务功能集合， $\mathcal{D} = \{1, \dots, d, \dots, D\}$  表示时延约束集合。

进一步，设  $\mathcal{P} = \{1, \dots, p, \dots, P\}$  表示设备处理速度集合， $\mathcal{M} = \{1, \dots, m, \dots, M\}$  表示设备的缓存队列长度集合， $\mathcal{F}' = \{1, \dots, f', \dots, F'\}$  表示设备具有不同功能。每个设备的业务调度遵循先到先服务(FCFS)的原则。

#### 3.2. 时延模型

为了刻画设备的实时处理能力，引入业务执行时延模型。该模型综合考虑了设备的处理性能、业务数据规模以及系统的排队特性，从而反映业务在设备端的整体响应特征。

令  $j_i$  表示为业务  $i$  的第  $j$  个子业务，其在设备  $u$  上的总时延由等待时延和处理时延两部分组成。子业务  $j_i$  的总时延如下：

$$T_{i,j} = n_{i,j} / p_u + W_{i,j,u} \quad (1)$$

其中， $n_{i,j}$  业务数据包大小， $p_u$  为设备  $u$  的处理能力， $W_{i,j,u}$  表示为根据排队论所得的平均等待时延。

#### 3.3. 问题公式化

多设备、多阶段的业务确定性调度问题旨在在调度过程中最大化满足时延约束的业务数量。该问题可形式化为一个集成业务、设备、时间与空间维度的四维匹配优化问题。其搜索空间随业务规模呈指数级增长，因此属于典型的 NP-hard 问题。

业务的完成时间必须满足其规定的时延约束，设业务  $i$  的完成时间为所有子业务中最后完成的子业务时间，即：

$$T_i^{end} = \sum_{j \in \mathcal{J}_i} T_{i,j}^{end} \quad (2)$$

其中， $\mathcal{J}_i$  为业务  $i$  的子业务集合， $T_{i,j}^{end}$  为子业务的完成时间。

那么业务  $i$  的时延约束表示如下：

$$T_i^{end} \leq d_i \quad (3)$$

此时，设一个二元变量  $x_i$ ，用于表示业务是否满足时延约束。如果满足，即  $x_i = 1$ 。反之，为 0。由于主要优化目标是满足时延约束业务数量最大化，多业务确定性调度优化模型被给出：

$$\max \sum_{i=1}^I x_i \quad (4)$$

## 4. 算法设计

本节通过将异构图神经网络(HGNN)与环境自适应深度强化学习(DRL)相结合，旨在高效求解业务调度问题。具体而言，首先将多业务、多设备、时间与空间耦合的四维匹配过程形式化为知识图谱。在此基础上，基于马尔可夫决策过程(MDP)对知识图谱中子业务与设备的实时状态及其状态转移进行动态建模，从而更准确、敏感地反映系统环境的变化。随后，引入熵正则化自适应策略，以在动态的网络环境中实现最优调度决策。

### 4.1. 知识图谱构建及 HGNN 嵌入

本节通过构建动态知识图谱来刻画多业务调度问题中复杂的依赖关系，如图 1 所示。本文采用知识图谱对业务调度过程中的状态转换与业务处理关系进行建模。具体而言，在业务时刻  $t$ ，系统状态可表示为一个异构图  $\mathcal{H}$ 。当在时刻  $t$  采取动作后，图中仅保留与该动作相关的依赖关系，并移除其他无关连接，从而生成下一时刻的图  $\mathcal{H}_{t+1}$ 。

随着动态知识图谱的持续更新，业务与设备之间的拓扑结构和语义关系也在不断变化。为了有效捕获这种多维异构关联，本文引入异构图神经网络对知识图谱中的节点与边进行表征学习。通过该方式能有效计算设备节点的嵌入向量  $\varphi_u'$ ，子业务节点的嵌入向量  $\psi_{i,j}'$  及系统在时刻  $t$  的全局状态向量  $h_t$ 。为后续的策略生成与调度决策提供高维、可微的状态输入。

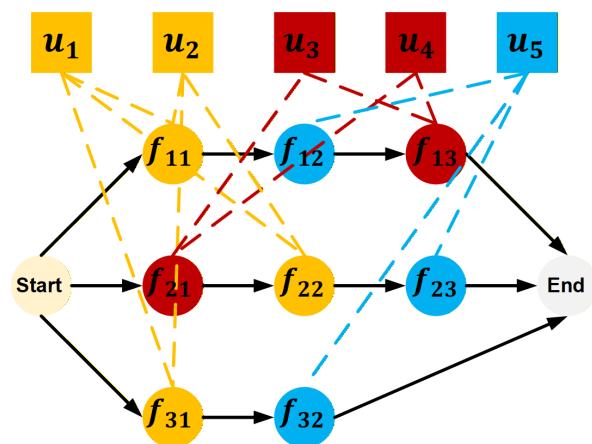


Figure 1. Knowledge graph  
图 1. 知识图谱

### 4.2. 基于知识图谱的马尔可夫决策

多业务调度过程本质上是一个由多个子业务持续分配至异构设备的动态映射过程，其状态演化仅依赖于当前的系统状态与决策行为。因此可建模为一个典型的马尔可夫决策过程。本文从 MDP 的视角将调度过程抽象为四个核心要素：状态、转移、动作和奖励，以形式化地描述子业务与设备间的动态映射及其最优决策生成过程。

**状态：**在每个决策时刻  $t$ ，系统状态  $S_t \in S$  用于表征业务与设备的全局调度特征，反映当前所有子业务的执行进度、设备的运行状态以及二者之间的依赖关系。形式化地，可将系统状态定义为：

$$S_t = \{S_t^{task}, S_t^{device}, S_t^{dep}\} \quad (4)$$

其中分别表示业务相关信息、设备相关属性和子业务与设备间的依赖关系，该依赖由动态知识图谱

$H_t$  进行结构化表示。

**动作:** 本文将决策动作  $a_t \in A_t$  定义为业务选择与设备分配的联合决策, 即在时刻  $t$  选择一个子业务 - 设备对。当子业务  $j_t$  可被调度且设备  $u$  的等待队列存在可用资源时, 将该子业务分配至设备  $u$  的业务队列中进行处理。

**转移:** 基于当前状态  $S_t$  和  $a_t$ , 环境转移到下一个状态  $S_{t+1}$ , 即设备队列情况发生变化时。

**奖励:** 对于单个业务而言, 本文关注其完成时间是否满足时延约束。对于整体的调度过程, 优化目标主要是满足时延约束的业务数量。因此对于多业务确定性调度问题的奖励函数设计如下:

$$r = \begin{cases} 1 + \alpha(1 - T/d), & T \leq d \\ -\beta[(T - d)/d], & T > d \end{cases} \quad (4)$$

式中的  $\alpha > 0$ , 鼓励业务更早完成。 $\beta > 0$ , 表示为轻度惩罚违约。将所有的  $r$  加起来即可得到所有业务的积累奖励。

### 4.3. 算法实现

本文使用熵系数自适应调整来增强策略的探索能力, 从而平衡探索与利用之间的关系。通过根据当前策略的熵值动态调整熵系数, 我们能够在训练过程中灵活应对不同环境的学习需求, 促进模型在复杂业务中的鲁棒性和适应性。在优化过程中, 我们的目标是动态调整熵系数以使策略的实际熵接近目标熵。当策略熵值大于目标熵值时, 降低熵系数以降低探索力度; 反之增加, 以增强探索。

## 5. 实验设计

### 5.1. 实验设置

本文设置 HGNN 迭代次数为 2, 设备和操作嵌入的维度为 16。对于训练, 将训练迭代次数和实例批大小设置为 5000 和 20。对于 PPO 损失函数, 将策略损失、价值损失和熵系数(动态调整)分别设为 1、0.5 和 0.01。PPO 优化周期设为 3, 折现系数设为 1。

### 5.2. 实验结果

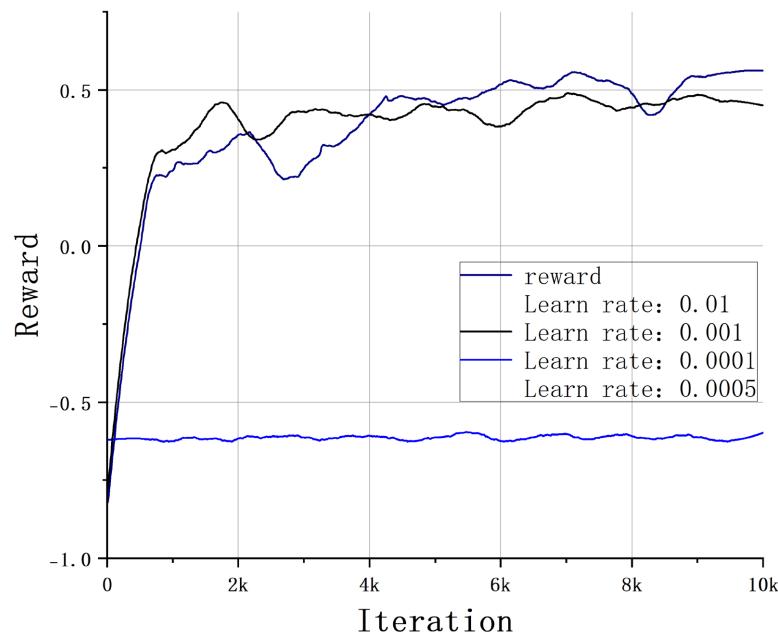
#### 5.2.1. 参数对比实验

在图 2 中, 我们研究学习率对算法性能的影响。从结果可以看出, 学习率为 0.0001 时表现出较为平稳且良好的收敛趋势。虽然初始阶段的收敛速度略慢于学习率为 0.001 和 0.0005 的情况, 但在训练后期, 0.0001 的曲线趋于平稳, 且最终取得了相对更好的长期表现, 避免了过大的波动和不稳定性。相比之下, 0.01 的学习率表现出明显的波动, 甚至没有显著的提升趋势。因此, 选择 0.0001 作为本实验的学习率, 有助于提高模型的稳定性和最终性能。

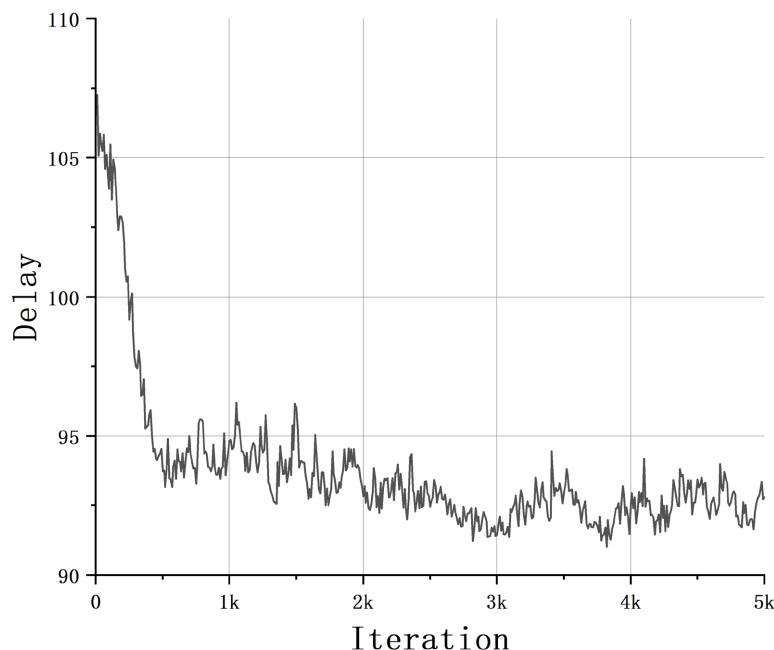
#### 5.2.2. 实验效果

从图 3 所示的性能指标可以看出, 业务完成的平均时延在整个实验过程中始终保持在较低水平, 并呈现出相对平稳的变化趋势。具体来看, 当业务请求量出现波动或不同类型任务同时到达时, 系统能够快速响应并有效安排任务执行, 避免了排队堆积和资源冲突导致的延迟增长。观察结果表明, 低且稳定的平均时延主要得益于调度策略在动态环境下对任务特性和系统资源状态的综合感知, 使得任务在多设备间得到合理分配, 从而减少了等待和执行开销。此外, 即使在业务类型高度异构、任务到达时间不均匀的情况下, 系统仍能维持整体处理效率, 这体现了调度策略在高并发电子商务场景下的鲁棒性和环境适应能力。综上所述, 本文方法不仅有效缩短了业务从生成到完成的整体响应时间, 也保证了系统在面

对复杂、多变的业务负载时仍能保持稳定的业务处理能力，为实现高效、可靠的电子商务任务调度提供了有力支持。



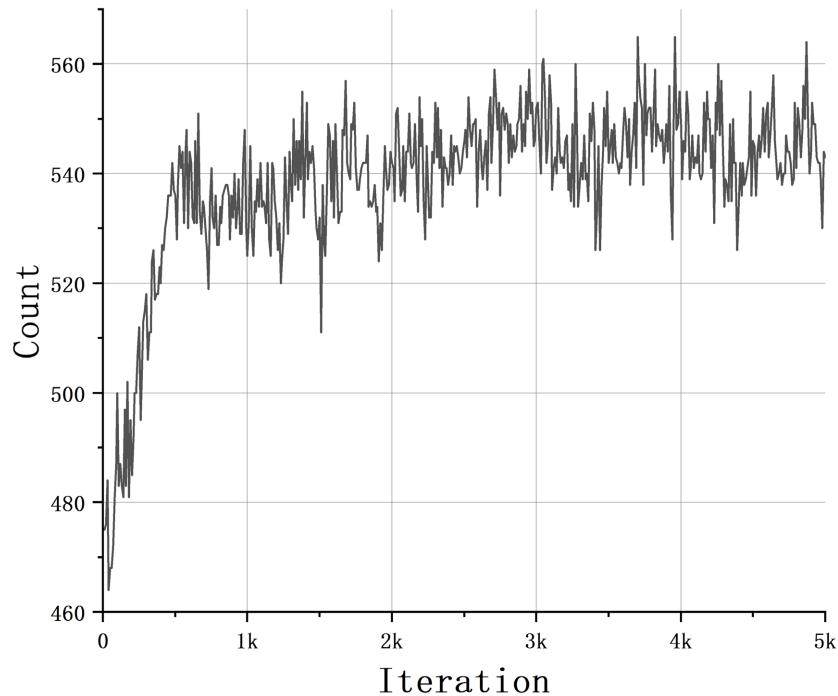
**Figure 2.** Learning rate comparison  
**图 2.** 学习率对比



**Figure 3.** Average task delay  
**图 3.** 业务的平均时延

进一步观察图4可知，在满足业务时延约束与服务级别协议(SLA)的前提下，本文方法实现的网络吞吐量始终维持在较高水平，并在整个实验过程中呈现出相对平稳的变化趋势。具体来看，当业务请求量

出现波动或不同类型任务同时到达时，吞吐量虽有一定波动，但整体仍保持稳定，说明系统能够高效调度多业务请求，减少任务在执行和排队过程中的资源冲突。实验数据还显示，在高并发和业务类型高度异构的情况下，系统吞吐量并未出现显著下降，表明调度策略能够动态适应业务负载变化，合理分配计算与通信资源，从而维持系统的高处理能力和稳定运行。此外，高且稳定的吞吐量反映了策略在处理复杂异构任务时的鲁棒性和环境适应性，为电子商务场景下多任务协同调度提供了有效的性能保证。



**Figure 4.** The number of tasks that meet the latency constraints  
**图 4.** 满足时延约束的业务数量

## 6. 结论

综上所述，本文针对电商业务在调度过程中呈现的差异化时间特性与资源利用效率不足的问题，提出了一种异构图神经网络增强的深度强化学习调度框架。通过对电商业务类别及业务-设备关联关系的显式建模，所提出的方法能够充分挖掘业务间依赖关系、设备资源状态以及业务类型约束，从而为调度决策提供更具表达力的结构化特征表示。同时，引入的熵正则化自适应策略在保证订单处理时延约束得到满足的前提下，有效提升了策略搜索的稳定性与调度性能。实验结果表明，该框架在复杂电商业务调度场景中能够实现高效、稳定的调度决策，为异构资源环境下的智能电商业务调度提供了一种可行且具有推广价值的解决方案。

## 参考文献

- [1] Kang, K., Ding, D., Xie, H., Yin, Q. and Zeng, J. (2022) Adaptive DRL-Based Task Scheduling for Energy-Efficient Cloud Computing. *IEEE Transactions on Network and Service Management*, **19**, 4948-4961. <https://doi.org/10.1109/tnsm.2021.3137926>
- [2] Tang, Z., Jia, W., Zhou, X., Yang, W. and You, Y. (2022) Representation and Reinforcement Learning for Task Scheduling in Edge Computing. *IEEE Transactions on Big Data*, **8**, 795-808. <https://doi.org/10.1109/tbdata.2020.2990558>
- [3] Jiang, F., Wang, K., Dong, L., Pan, C. and Yang, K. (2020) Stacked Autoencoder-Based Deep Reinforcement Learning

- for Online Resource Scheduling in Large-Scale MEC Networks. *IEEE Internet of Things Journal*, **7**, 9278-9290. <https://doi.org/10.1109/iot.2020.2988457>
- [4] Orr, M. and Sinnen, O. (2015) A Duplicate-Free State-Space Model for Optimal Task Scheduling. In: Träff, J.L., *et al.*, Eds., *Euro-Par 2015: Parallel Processing*, Springer, 97-108. [https://doi.org/10.1007/978-3-662-48096-0\\_8](https://doi.org/10.1007/978-3-662-48096-0_8)
- [5] Zheng, J., Okamura, H. and Dohi, T. (2023). Hierarchical Dependability Modeling with Multi-State Systems. 2023 IEEE 28th Pacific Rim International Symposium on Dependable Computing (PRDC), Singapore, 24-27 October 2023, 268-277. <https://doi.org/10.1109/prdc59308.2023.00043>
- [6] Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C. and Yu, P.S. (2021) A Comprehensive Survey on Graph Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems*, **32**, 4-24. <https://doi.org/10.1109/tnnls.2020.2978386>
- [7] Liu, C.-L. and Huang, T.-H. (2023) Dynamic Job-Shop Scheduling Problems Using Graph Neural Network and Deep Reinforcement Learning. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, **53**, 6836-6848. <https://doi.org/10.1109/tsmc.2023.3287655>
- [8] Peng, Y., Guo, J. and Yang, C. (2024) Learning Resource Allocation Policy: Vertex-GNN or Edge-GNN? *IEEE Transactions on Machine Learning in Communications and Networking*, **2**, 190-209. <https://doi.org/10.1109/tmfcn.2024.3354872>
- [9] 覃雪莲, 刘志学. 考虑供应商竞争的电子商务供应链物流服务模式优化研究[J]. 中国管理科学, 2024, 32(1): 125-136.
- [10] 杜艳鑫. 基于 AI 的电商优化服务模式探索[J]. 电子商务评论, 2025, 14(7): 1662-1670.