

基于深度强化学习算法的弹性供应链调度优化方法

张进军¹

ZHANG Jinjun

摘要

由于参与供应链的客户需求和供应者配置的多元性,导致供应链的配送成本难以得到有效控制,主要是因为模型求解过程中,受困于参数本身的矛盾性,求解过程很容易陷入局部最优。为此,提出基于深度强化学习算法的弹性供应链调度优化方法研究。分别从供应者配置角度和客户需求角度构建了供应链模型,以供应链配送成本最小化为目标函数,应用深度强化学习算法中的深度Q网络(deep Q-network, DQN)算法进行训练,同步进行弹性供应链优化调度。DQN能够有效地处理这种高维状态空间,通过深度神经网络学习状态与动作之间的映射关系,自动提取关键特征,从而简化问题的复杂性。将收敛输出结果期望误差,输入供应链模型进行迭代计算,输出优化调度结果。测试结果表明,设计的方法可以实现对配送成本的有效控制。

关键词

深度强化学习算法;弹性供应链调度;供应者配置;客户需求;供应链模型;配送成本最小化;Deep Q-Network;误差收敛

doi: 10.3969/j.issn.1672-9528.2024.04.019

0 引言

弹性供应链是指在面对外部环境不确定性时,能够快速、有效地调整供应链的运作,以适应变化并保持稳定输出的供应链系统^[1]。这种供应链管理强调是对市场变化的快速响应和适应性,而非传统的以生产成本和效率为主的供应链管理^[2]。在弹性供应链中,供应者和客户需求的不确定性影响着供应效率。供应者的不确定性主要来自供应商的供货能力、产品质量、交货时间等方面的不确定性^[3]。这种不确定性可能导致供应链中断,影响生产计划的执行。客户需求的不确定性是另一个重要的不确定因素。客户需求的变化可能导致生产计划调整、库存积压、订单取消等一系列问题^[4]。因此,有必要研究一种有效应对供应者和客户需求的不确定性的方法,建立灵活的生产调度方法,降低不确定性带来的风险,提升竞争力。

文献[5]使用禁忌搜索算法中的禁忌表机制对搜索过程进行限制,结合离散化实数编码方式,将连续的调度问题转换为离散的编码问题,引入改进哈里斯鹰算法不断迭代优化,求解最优值。但改进哈里斯鹰算法对初始解的依赖性较强,不同的初始解导致搜索结果差异较大。文献[6]基于供应链

协同物流配送机制,建立考虑多车型的物流配送路径调度数学模型,并以物流配送总成本为目标函数,通过改进粒子群算法寻找最佳物流配送路径调度方案。该算法中对物流配送总成本的优化无法充分考虑客户满意度,可适用空间较小。

为此,本文提出基于深度强化学习算法的弹性供应链调度优化方法。以供应者和客户两方面为重点考量,将深度强化学习算法应用于弹性供应链的调度优化中。通过建立深度神经网络模型,根据供应者和客户两方面的实时数据,进行自动学习,并根据实时供应者和客户数据输出动态优化调度决策。

1 弹性供应链调度优化方法设计

1.1 供应链模型构建

供应链主要由供应者和客户两部分组成,从供应者配置的角度进行分析,产品质量、交货期、价格以及具体的服务质量是影响供应链配送执行的主要因素^[7],因此,可以将其表示为:

$$T_{TC} = \sum_{s=1}^S w_s (\sum_{i=1}^I \sum_{j=1}^J c_{ij} q_{ij} + \sum_{i=1}^I \sum_{j=1}^J C_{ij} Q_{ij}) \quad (1)$$

式中: T_{TC} 表示供应链中供应者的配置成本; s 表示参与供应链的供应者数量; w_s 表示供应者的单位配送路径总长度,即配送距离; c_{ij} 表示交货期约束下的供应者 i 与客户 j 之间的单位配送成本; q_{ij} 表示交货期约束下的供应者 i 与客户 j 之间的配送时间开销; C_{ij} 表示超时状态下供应者 i 与客户 j 之

1. 安徽警官职业学院信息管理学系 安徽合肥 230031

[基金项目] 2023年度安徽省高校自然科学研究重大项目“基于元知识的弹性供应链优化模型的设计”(2023AH040354)

间的单位配送成本； Q_{ij} 表示超时状态下供应者 i 与客户 j 之间的配送时间开销。

供应链风险主要为供应链配送超时^[8]，对其进行量化时，其计算方式表示为：

$$T_{PC} = \sum_{s=1}^S w_s \sum p c_j (1 + \lambda z_j) \quad (2)$$

式中： T_{PC} 表示供应链风险系数； p 表示风险因素发生的概率参数； c_j 表示客户 j 导向下的单位配送成本变动系数； λ 表示物流路径和运输方式受风险因素的扰动系数； z_j 表示客户 j 导向下的中断事件对供应链产品生产成本的影响程度。

从客户需求的角度进行分析，客户的需求主要体现为产品需求量、客户需求位置两部分。其中，客户需求位置决定了供应链的配送距离，产品需求量决定了配送执行的最小单位^[9]。因此，其可以表示为：

$$T_{UC} = \sum_{u=1}^U k_u \sum c_k u_{ks} \quad (3)$$

式中： T_{UC} 表示供应链中客户的需求成本； u 表示参与供应链的客户数量； k_u 表示客户可执行的单位配送路径总长度，即最小配送距离； c_k 表示客户可执行的单位配送成本； u_{ks} 表示客户需求导向下的产品需求量。

按照上述所示的方式，实现对供应链中供应者和客户状态的分析，那么此时的供应链模型就可以表示为客户需求导向下，供应者可执行产品总量的适应性匹配，即：

$$F = \sum T_{UC} + T_{TC} + T_{PC} \quad (4)$$

式中： F 表示供应链总成本模型， t 表示供应者风险值。

按照上述所示的方式，实现对供应链模型的构建，确定了供应链的总成本，为后续的弹性供应链调度优化提供可靠的执行基础。

1.2 基于深度强化学习算法的弹性供应链调度

结合 1.1 部分构建的供应链模型，本文在具体的弹性供应链调度过程中，引入了深度强化学习算法^[10]，并以供应链配送成本最小化为目标函数，对具体的供应链配送资源配置进行迭代优化。基于此，目标函数表示为 $\min F$ 。

本文应用了 Deep Q-Network 算法进行弹性供应链优化调度。供应链调度问题通常具有高维状态空间，因为需要考虑众多因素如库存状态、订单量、交货时间等。DQN 能够有效地处理这种高维状态空间，通过深度神经网络学习状态与动作之间的映射关系，自动提取关键特征，从而简化问题的复杂性。且 DQN 通过不断与环境交互进行学习，能够实时地调整和优化调度决策，可以有效适应供应链环境经常变化的环境，快速响应市场变化和客户需求。

首先确定了神经网络的输入层，以供应链环境的状态变量作为输入数据，包括当前来自客户的订单数量、供应者

的当前库存水平和生产状态、可执行的物流服务状态。本文将运输车辆的位置和可用性作为物流服务状态的具体参数信息。

本文将弹性供应链调度问题的复杂性和数据特性作为隐藏层的设计基准，对神经网络的隐藏层进行设计。考虑到实际的供应链状态变量为序列数据，因此构建了以全连接层为基础的隐藏层。具体可以表示为：

$$C_p = \max \gamma(1 - \beta) \quad (5)$$

式中： C_p 表示神经网络的隐藏层， γ 表示全连接层连接系数， β 表示隐藏层的映射函数。

基于 1.1 节构建的供应链模型，体现了弹性供应链调度涉及的参数是多元的，因此，按照供应者和客户两个角度，设计隐藏层的数量设置为 2 层，每个隐藏层的具体神经元数量以供应者和客户的具体参数配置种类为基础进行差异化设置。在建立二者之间的加权连接系数时，利用风险状态进行设置。

最后设计神经网络输出层，设置网络期望的输出为一种调度方案（即 Q 值）。使用梯度下降算法更新神经网络的权重和偏置，以最小化预测 Q 值与实际 Q 值之间的差异。定义损失函数：

$$L = (Q_{\text{predicted}} - Q_{\text{actual}})^2 \quad (6)$$

式中： $Q_{\text{predicted}}$ 表示神经网络预测的 Q 值， Q_{actual} 表示实际的 Q 值。

然后，对损失函数求关于权重 W 和偏置 b 的梯度，得到损失函数对应的梯度向量：

$$\begin{aligned} \nabla L_w &= \frac{\partial L}{\partial W} \\ \nabla L_b &= \frac{\partial L}{\partial b} \end{aligned} \quad (7)$$

接下来，使用梯度下降算法更新权重和偏置，即将当前的权重 W 和偏置 b 分别减去学习率与对应的梯度相乘得到的变化量：

$$\begin{aligned} W' &= W - \lambda \nabla L_w \\ b' &= b - \lambda \nabla L_b \end{aligned} \quad (8)$$

式中： λ 为学习率，用于控制每次更新的步长。

在神经网络训练中，使用反向传播算法计算这些梯度值。通过链式法则，将损失函数的梯度向后传播，以确定每层神经元对误差的贡献，并最终计算出对应权重 W 和偏置 b 的梯度。通过反复迭代以上步骤，不断更新神经网络的权重和偏置，使得损失函数逐渐减小，从而使预测的 Q 值与实际 Q 值之间的差异越来越小，获得更好的预测结果。

当网络进行一次动作并收到奖励时，利用使用经验回放（experience replay）技术将对应的供应链状态以及调度动作存储在经验回放缓冲区中。需要注意的是，在训练时，在

缓冲区中随机选择采样经验时，可使用最近的经验，以此增加模型的稳定性和探索能力。当构建的 Deep Q-Network 输出与期望输出的误差达到收敛状态后，使用训练好的 Deep Q-Network 生成调度决策，以此结合输出的参数信息，调整分配供应链的订单供应者和调度运输执行目标。

按照上述所示的方式，根据实时供应链的状态，调整调度供应链决策，以适应市场需求的变化和供应链的动态特性。

2 测试与分析

2.1 测试准备

在对本文设计的基于深度强化学习算法的弹性供应链调度优化方法实际应用效果进行分析时，以某实际的由电子产品的多个零件供应商、单个成品（智能手机）制造商构成的供应链为基础，在 MATLAB 仿真平台上开展了对比测试。设置算法的初始学习率为 0.5，最终学习率为 1e-3，货损风险为 0.18，折扣因子为 0.99，衰减率 0.999，网络更新频率 1000 步，最大训练步数 10 000 步。

对测试供应链的具体情况进行分析，其是一组服务于特定顾客的分配中心。在每个供应周期内，对应的供应产品由制造商完成生产后，需要运往配送中心开展后续的销售等活动。其中，在测试供应链中，对应的制造者生产单周期能力限制为 4000 件。具体参与供应链的 9 个供应者中，相关供应能力、供应柔性、订货成本、中断概率以及运送时间信息如表 1 所示。

表 1 测试供应链供应者参数信息统计表

供应者编号	最大供应量 / 件	供应柔性系数	单位采购成本 / 元	中断概率 / %	运送时间 / h
G001	4500	0.30	400	1.25	26
G002	4500	0.45	450	1.33	26
G003	4500	0.26	600	2.45	28
G004	4500	0.33	600	2.02	28
G005	4500	0.35	550	1.78	30
G006	4500	0.40	600	1.56	30
G007	4500	0.44	520	1.50	30
G008	4500	0.45	520	3.35	32
G009	4500	0.60	500	4.12	26

结合表 1 所示的信息可以看出，在参与供应链的 9 个供应者中，对应的最大供应量均为 4500 件，但是具体的供应时间以及相关成本和柔性系数均存在不同程度的差异，这也使得其实际的中断概率表现出了不同的特点。在具体的供应链调度过程中，需要结合实际情况进行适应性调整。

统计供应链中的 15 个客户信息，其中，相关需求量、惩罚成本、交期等参数信息如表 2 所示。

表 2 测试供应链客户参数信息统计表

分配中心	客户编号	需求量 / 件	单位拖延惩罚成本 / 元	未完成惩罚成本 / 元	供应链运输时间 / h
S1	K001	500	11	35	24
	K002	700	11	26	24
	K003	850	12	24	24
	K004	1200	12	25	24
	K005	1200	13	25	24
S2	K006	1100	11	26	28
	K007	900	12	19	28
	K008	950	12	19	28
	K009	1100	13	20	28
	K010	1100	10	21	28
S3	K011	850	12	22	30
	K012	700	11	22	30
	K013	700	10	23	30
	K014	850	11	23	30
	K015	900	12	25	30

结合表 2 所示的信息可以看出，参与供应链的客户主要分布在三个配送中心周围，对应的需求量以及最晚配送周期各有不同，不同区域的运输时间分别为 24 h、28 h 和 30 h，不同客户拖延与未完成运输订单的惩罚成本也有所不同。

以上述测试环境为基础，分别采用本文设计的基于深度强化学习算法的弹性供应链调度优化方法，以及文献 [5] 方法和文献 [6] 方法作为测试的对照组，测试不同调度方法下的供应链运输效果。

2.2 测试结果与分析

2.2.1 运输成本

在对不同调度方法下的测试供应链运输效果进行分析时，本文将不同配送中心对应的配送成本作为具体的评价指标，得到的测试结果如图 1 所示。

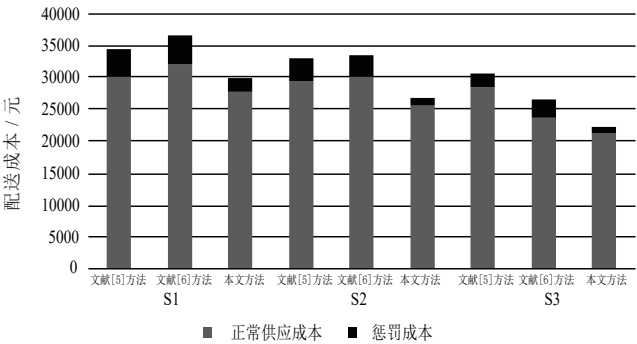


图 1 不同调度方法运输成本测试结果对比图

分析图 1 可知，在三种不同调度方法的测试结果中，对于不同配送中心对应客户的供应链配送成本表现出了较为明

显的差异。在三种配送中心对应的应客户供应链配送成本中, 本文设计调度方法下的正常供应成本和惩罚成本均低于其他方法, 从整体角度看, 本文设计调度方法下的供应链配送成本为 78 994 元, 分别低于文献 [2] 调度方法 19 310 元, 低于文献 [3] 调度方法 17 715 元。这表明本文方法可以实现对配送成本的有效控制, 具有良好的实际应用价值。

2.2.2 配送效率

基于上述实验结果, 进一步验证本文设计调度方法下的供应链调度后的配送效率, 验证在供应链中, 配送至各客户站点所需运输时间, 结果如图 2 所示。

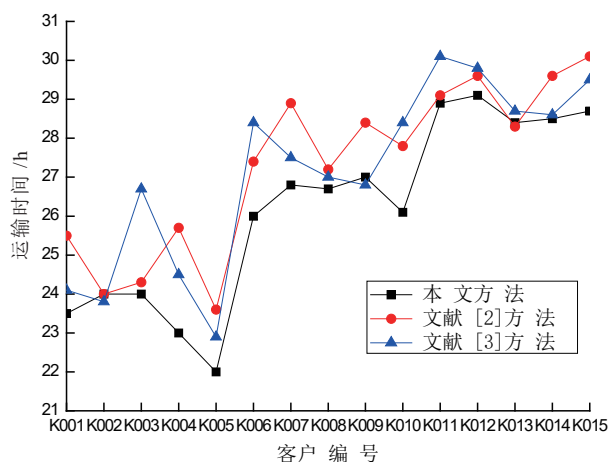


图 2 配送效率

分析图 2 可知, 基于三种不同的方法进行供应链调度后, 各客户站点配送时间存在较大的差别。显然, 本文设计的调度方法下可有效节约到达各客户站点的时间, 从而提升了整体的供应链配送效率, 使得配送时间保持在要求配送时间以内最少可以节约 0.48 h, 而文献 [2] 和文献 [3] 方法的供应链配送效率明显低于本文方法, 在部分客户站点配送中, 甚至出现超出要求配送时间的情况, 这表明本文方法具有更为优越的配送效率。

综上所述, 本文基于 Deep Q-Network 算法的调度方法, 在供应者配置和用户需求方面的考虑能够提高供应链的配送效率、降低成本, 并具有良好的实时性和可扩展性。这是因为本文设计的调度方法考虑了不同配送中心的供应者配置。通过 Deep Q-Network 算法, 可以根据实际情况确定每个配送中心所需的供应者数量和配置。这种个性化的供应者配置能够更好地适应不同的需求, 提高配送效率和成本控制。另外, 本文方法还考虑了用户的需求。通过 Deep Q-Network 算法, 根据不同的环境和任务进行自适应调整。这使得调度方法在面对不断变化的供应链需求时, 能够快速优化和调整策略, 从而有效降低了惩罚成本。

3 结语

本文提出基于深度强化学习算法的弹性供应链调度优化方法, 充分考虑了供应者与客户之间的关系以及具体的配置和需求情况, 借助深度强化学习算法切实实现了成本导向下的有效调度, 降低了相同供应链配送任务下的成本投入。本文对于弹性供应链调度优化问题的分析与研究, 可以为实际的供应链调度工作提供有价值的参考。

参考文献:

- [1] 彭小剑, 程兆兆, 谢梦萍, 等. 新零售时代下“智慧物流与新型供应链”协调发展研究[J]. 中国航务周刊, 2023(47): 182-184.
- [2] 舒双武, 房芳, 周俊伟, 等. 拉式供应链模式下卷烟配送物流资源的优化重构和动态适配研究与应用[J]. 物流技术与应用, 2023,28(11):115-119.
- [3] 王金刚, 王树海, 马长璐, 等. 大型能源集团一体化煤炭供应链实时调度智能决策关键技术研究[J]. 能源与环境, 2023, 45(10):264-272.
- [4] 赵楠, 王俊, 黄桦, 等. 计及氢电混合动力车响应的多区域综合能源系统协调优化调度[J]. 电力自动化设备, 2023(12): 262-269.
- [5] 支永坤, 刘欢, 代永强. 改进的哈里斯鹰算法求解农产品供应链调度[J]. 计算机应用研究, 2023,40(2):413-417+449
- [6] 赵静, 唐业富. 考虑多车型的供应链运输路径协同调度算法[J]. 计算机仿真, 2023,40(3):181-185
- [7] 曹仟妮, 贾孟硕, 李博达, 等. 面向盐穴大规模储氢商业模式的副产氢供应链管理决策[J]. 清华大学学报(自然科学版), 2023,63(12):2019-2032.
- [8] 欧阳世波, 张森寒, 谢俊明, 等. 计及多目标约束的成品卷烟销售订单配送调度优化[J]. 中国烟草学报, 2023,3(27):1-14.
- [9] 唐红涛, 沈毅, 张伟, 等. 改进鲸鱼算法求解分布式装配柔性作业车间生产与配送联合调度问题[J]. 计算机应用研究, 2023, 40(7):1982-1990.
- [10] 赵丹, 张璨辉, 刘翔斌, 等. 基于遗传算法的计量资产供应链均衡调度模型[J]. 自动化技术与应用, 2023,42(1):152-154+162.

【作者简介】

张进军 (1979—), 男, 安徽泗县人, 本科, 讲师, 研究方向: 计算机教育、数据挖掘、数据分析。

(收稿日期: 2024-01-09)