

基于学习型算法的农资配送优化问题研究

张吉哲¹, 刘欢¹, 代永强¹, 秦立静¹, 支永坤²

(1. 甘肃农业大学信息科学技术学院, 甘肃兰州 730070; 2. 兰州信息科技学院信息工程学院, 甘肃兰州 730300)

摘要: 针对农资供应链订单配送路径优化问题, 考虑新能源货车综合续航里程、车辆最大载荷能力和时间窗等约束, 建立车辆路径规划问题的数学模型(DCVRPTW)综合优化车辆固定成本和运输成本, 提出一种基于深度强化学习的群智能优化算法框架(DRL-SIA)。智能体就是决策者, 以环境状态为输入选出动作池中最佳动作改变环境并获得环境奖励。DRL-SIA算法结合训练后的智能体与群智能算法以代替原算法进行决策选择, 从而提升寻优速度与精度。实验表明, 所提算法的最优解相较于其他算法在所有算例中最优, 验证了该算法能有效降低农用物资供应链中的物流运输成本。

关键词: 深度强化学习; 车辆路径规划; 群智能优化算法; 农资供应链; 演化计算

DOI: 10.11907/rjdk.231948

开放科学(资源服务)标识码(OSID):

中图分类号: TP18

文献标识码: A

文章编号: 1672-7800(2024)009-0122-09



Research on Optimization Problem of Agricultural Supplies Distribution Based on Learning Algorithm

ZHANG Jizhe¹, LIU Huan¹, DAI Yongqiang¹, QIN Lijing¹, ZHI Yongkun²

(1. College of Information Science and Technology, Gansu Agricultural University, Lanzhou 730070, China;

2. College of Information Engineering, Lanzhou Institute of Information Technology, Lanzhou 730300, China)

Abstract: A mathematical model for vehicle path planning (DCVRPTW) was established to optimize the delivery path of agricultural inputs supply chain orders, taking into account constraints such as the comprehensive range, maximum load capacity, and time window of new energy trucks. The model comprehensively optimizes the fixed and transportation costs of vehicles, and proposes a swarm intelligence optimization algorithm framework based on deep reinforcement learning (DRL-SIA). An intelligent agent is a decision-maker who selects the best action from the action pool based on the environmental state as input to change the environment and obtain environmental rewards. The DRL-SIA algorithm combines trained agents with swarm intelligence algorithms to replace the original algorithm for decision selection, thereby improving optimization speed and accuracy. The experiment shows that the optimal solution of the proposed algorithm is superior to other algorithms in all cases, verifying that the algorithm can effectively reduce logistics transportation costs in the agricultural material supply chain.

Key Words: deep reinforcement learning; vehicle routing planning; swarm intelligence optimization algorithm; agricultural supply chain; evolutionary computation

0 引言

农资是保障我国农业发展的重要物资, 根据《第三次全国国土调查主要数据公报》数据显示, 目前全国耕地面积为191 792.79万亩, 2022年粮食总产量达到68 653万吨^[1]。这些数字的背后是对农资的极大需求, 例如化肥每

年需要5 191.26万吨、农药每年需要139.17万吨^[2]。由于农资产品具有毒性、易潮解、季节性强、体积大、重量大、储存条件严苛等特性, 物流服务质量标准相较于其他产品标准更高。然而, 农资消费地大部分分布在物流设施不完善的农村等偏远地区, 配送难度更大, 因此农用物资电子商务平台需要搭建农资供应链, 做好农资存储与平台仓库配送到消费者手中的物流^[3]。

收稿日期: 2023-09-06

扫描二维码阅读全文:



基金项目: 甘肃省自然科学基金项目(21JR7RA204, 1506RJZA007); 甘肃省高等学校教育创新基金项目(2022B-107, 2019A-056)

作者简介: 张吉哲(1997-), 男, 甘肃农业大学信息科学技术学院硕士研究生, 研究方向为演化计算; 刘欢(1990-), 男, 硕士, 甘肃农业大学信息科学技术学院讲师, 研究方向为演化计算。本文通讯作者: 刘欢。

现代农业是与工业4.0或后工业时代对称的农业现代化,不同于农业产业化与农业工业化,是以智慧经济为主导、大健康产业为核心的自动化、个性化、艺术化、生态化、规模化、精准化的智慧农业,需要将所需农资产品准时配送到消费者的田间地头,保障农资高效使用,减少在消费者手中因仓储不当导致的功效损失^[4]。为落实中央提出的“双碳”政策,对农资供应链质量提出了更高的要求,例如新搭建的物流体系通常使用新能源货车^[5]。

为了使农资配送成本效益更优,研究学者采取了不同优化目标,例如减少配送距离、时间、成本、碳排放等。基础研究是将一个区域内的农资配送问题设计为车辆路径问题(Vehicle Routing Problem, VRP),并考虑一些限制因素,最常见的约束就是车辆载重约束,即可将问题转化为有容量约束的车辆路径调度问题(Capacitated Vehicle Routing Problem, CVRP)^[6,7]。国内外学者针对 CVRP 在模型构建、实际生产约束的提炼、调度方案的求解算法等方面作了很多探索及研究,但由于距离和能力约束的车辆调度问题(Distance Capacitated Vehicle Routing Problem, DCVRP)的复杂性及特殊性,导致最常用的车辆路径规划问题研究成果相对较少,尤其是为了保障农资准时送达,加入满足收货点时间窗约束的 DCVRP 问题研究更少。因此,DCVRP 模型应考虑车辆行驶距离、车辆载重、收货点可收货时间窗等多方面的实际生产约束。

为此,不少研究者提出了相应的群智能优化算法、多智能体方法及混合方法。例如, Borcinova^[8]将原问题分解为更容易求解的子问题,在此基础上借鉴内核搜索算法思想提出一种解决 CVRP 的数学方法。Mahmood^[9]提出一种猫鼬氏族算法(Mongoose Clan Algorithm, MCA)求解 CVRP 问题,根据猫鼬行为将解决方案划分为子组,运用分而治之的思想为寻找最佳解决方案提供了广泛的选择。Birecik 等^[10]提出一种交互式模糊方法解决每辆车和需求行程时间不精确的 DCVRP 问题。Cai 等^[11]提出一种称为动态空间缩减蚁群优化(Dynamic Space Reduction Ant Colony Optimization, DSRACO)的新型蚁群算法(Ant Colony Optimization, ACO)解决有能力的车辆路径问题。Dalbah 等^[12]提出一种冠状病毒群体免疫优化器(Coronavirus Herd Immune Optimizer, CHIO)解决有能力的车辆路径问题。Rattanamanee 等^[13]使用贪婪交换和 2-opt 改进算法进行多次迭代,在合理的计算时间内获得接近最优多行程车辆路径问题(Multi-Trip Vehicle Routing Problem, MTVRP)的解决方案。

针对有时间窗口的距离和能力约束的车辆调度模型(Distance Capacitated Vehicle Routing Problem with Time Windows, DCVRPTW)的求解差异性、问题空间较复杂等问题,通过深度强化学习构建概率模型挖掘 DCVRPTW 模型的结构特性,在算法训练过程中隐形使用这些信息。由于深度强化学习算法在函数优化问题及混合流水车间调

度问题(Hybrid Flow Shop Scheduling Problem, HFSP)^[14]、布尔可满足性问题(SAT)^[15]等组合优化问题应用效果较好,在各方向均具有实际研究价值。因此,本文基于深度强化学习算法的自主学习能力探索有潜力的问题空间,使用深度强化学习算法框架结合 DCVRPTW 问题模型(DRL-SIA)解决实际农资配送问题。将深度强化学习的自适应决策能力与群智能算法的寻优能力相结合,智能体通过训练可有效指导群智能算法决策,使群智能算法根据问题特性寻优以节省行驶距离和车辆数量,从而达到减少人工成本、燃料成本、交货时间和温室气体排放的目的。

在 DRL-SIA 中,首先通过训练 DQNAgent 模块对候选解进行寻优操作,然后利用奖惩函数优化 DQNAgent 网络参数分析 DCVRPTW 问题特性以提升配送中心车辆利用率,最后对车辆路径方案进行进一步强化,仿真实验与统计分析证明了 DRL-SIA 求解有距离和能力约束带时间窗的车辆路径规划问题的可行性与稳定性。

1 问题模型

1.1 模型描述

本文研究问题模型来源于农资电商平台的农资订单配送,配送过程可抽象为配送中心、收货点、配送车辆。在配送中心有多辆可同时使用的车辆,收货点存在时间窗限制、车辆存在载重和能源限制。因此,将其路径规划问题抽象为 DCVRPTW,如图 1 所示。假设 DCVRPTW 中共有 m 个收货点,配送中心出发的车辆最终返回配送中心,配送中心车辆参数一致,但收货点时间窗口不同。

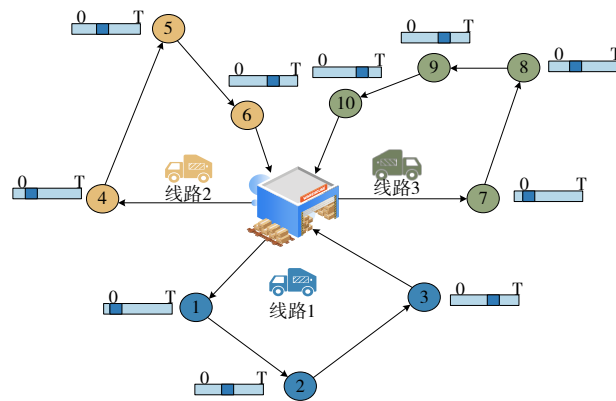


Fig. 1 DCVRPTW system model

图 1 DCVRPTW 系统模型

由于合理的假设可简化模型施工过程,减少二级因素的影响和模型构建的难度。为此,本文为了控制一些无法解释的变量,提出模型构建所需的环境设置条件:①收货点满足每个收货点需求,只允许一个配送运输车辆为其服务且仅访问一次,除此之外任意收货点需求量还需小于车辆最大载重;②配送车辆类型统一,具有相同的负载和速度;③配送路线根据需求设定,配送中心有足够车辆在各路线上运输,运输为配送中心到收货点的单向运输;④每

个配送车辆的开始和最终目的地为配送中心,不再进行中途分配;⑤同一路径上所有客户需求的总量不得超过车辆额定载重;⑥每个配送车辆均需返回配送中心补能,单程距离不能超过最大行驶距离;⑦时间窗口至少满足车辆路径只包含一个需求节点的情况,交付车辆到达时间和装载操作应在交付点可接受时间窗口内满足;⑧配送过程不考虑道路拥堵、恶劣天气状况或车辆故障等情况。

1.2 数学模型

表1为本文问题模型中数学符号的意义。在DCVRPTW中主要优化运输成本,具体的混合整数规划模型如式(1)一式(5)所示。

Table 1 Mathematics symbol

表1 数学符号

符号	实际意义	符号	实际意义
n	车辆数量	m	配送点数量
i, j, q	配送点编号	k	车辆编号 $k \in \{1, 2, \dots, n\}$
f_1	每辆车使用固定成本	p	配送中心编号
D_{ij}	配送点 i 和 j 之间的距离	f_2	每辆车单位距离使用成本
X_{ijk}	0, 1 决策变量, 表示第 k 辆车为配送点 i, j 完成交付操作	X_{ik}	0, 1 决策变量, 表示第 k 辆车为配送点 i 完成交付操作
Q_k	第 k 辆车的实际载重	X_{ipk}	0, 1 决策变量, 表示第 k 辆车为配送中心 p 和配送点 i 完成交付操作
Q	配送车辆最大载重	Q_i	配送点 i 的需求重量
e_p	配送中心 p 最早送货时间	e_i	配送点 i 最早接货时间
l_p	配送中心 p 最晚送货时间	l_i	配送点 i 最晚接货时间
A_{pk}	配送中心 p 实际送货时间	A_{ik}	配送点 i 实际接货时间

目标函数:

$$\text{Min}C = C_1 + C_2 \quad (1)$$

$$C_1 = n \times f_1 \quad (2)$$

$$C_2 = \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^n D_{ij} \times f_2 \times X_{ijk} \quad (3)$$

式中: C_1 为固定成本; C_2 为运输成本。

约束条件:

$$X_{ijk} = \begin{cases} 1 & \text{表示车辆 } k \text{ 从 } i \text{ 行驶到 } j \\ 0 & \text{表示车辆 } k \text{ 不从 } i \text{ 行驶到 } j \end{cases} \quad (4)$$

$$X_{ik} = \begin{cases} 1 & \text{表示车辆 } k \text{ 为 } i \text{ 服务} \\ 0 & \text{表示车辆 } k \text{ 不为 } i \text{ 服务} \end{cases} \quad (5)$$

$$\sum_{i=1}^m \sum_{k=1}^n X_{ipk} = \sum_{i=1}^m \sum_{k=1}^n X_{pik} \quad (6)$$

$$\sum_{i=1}^m X_{ijk} = \sum_{i=1}^m X_{jik} \quad (k \in n, \forall j \in m) \quad (7)$$

$$Q_k \leq Q \quad (8)$$

$$\sum_{k=1}^n X_{ik} = 1 \quad (\forall i \in m) \quad (9)$$

$$Q \geq Q_i \quad (\forall i \in m) \quad (10)$$

$$\sum_{i=1}^m X_{iqk} = \sum_{j=1}^m X_{jqk} \quad (k \in n, \forall q \in m) \quad (11)$$

$$e_i \leq A_{ik} \leq l_i \quad (12)$$

$$e_p \leq A_{pk} \leq l_p \quad (13)$$

$$L_k = \sum_{i=1}^m X_{ijk} \times D_{ij} \quad (k \in n, \forall j \in m) \quad (14)$$

$$L_k \leq L \quad (15)$$

式(1)为DCVRPTW的目标函数,即在满足配送时效的基础上实现配送成本最小化;式(2)为与车辆行进的距离及与运输货物数量无关的费用,包括管理费用、维修和保养费用、道路养护费、桥梁费和其他杂项费用,这部分成本只与车辆数量有关;式(3)表示在分配过程中随着里程和车辆负载增加或减少的部分成本,也被称为车辆公里可变成本,包括能源消耗和折旧成本,可将配送车辆的里程和每单位距离的运输成本相乘获得;式(4)、式(5)都通过0-1变量表示是否经过决策;式(6)表示所有车均由配送中心出发,完成所有配送任务后返回配送中心;式(7)表示每个收货点进出车辆相等;式(8)表示每辆车载重不能超过其最大载重限制;式(9)表示每个收货点仅允许一个配送运输车辆为其服务且访问一次;式(10)表示车辆最大载重满足任意收货点的需求量;式(11)表示一个送货任务完成后送货车辆必须前往下一个送货点;式(12)表示满足所有收货点的时间窗约束;式(13)表示满足所有配送中心的时间窗约束;式(14)、式(15)表示每辆车的行驶距离不能超过最大行驶距离。

在DCVRPTW中存在以下两个子问题:①车辆路径如何切割;②如何满足众多限制条件。车辆路径如何切割决定了是否能满足限制条件,满足所有限制的路线决定了配送成本,因此应协同考虑这些子问题,将问题特征融合在DRL-SIA框架中才能获得质量较高的候选解。实际中,当农用物资电子商务平台中心仓库搭建完毕后,平台销售订单信息会进入仓库信息库,仓库需要在要求时间内将货品配送到收货地,因此DRL-SIA框架需要对农资配送路径进行规划,以最低成本完成农资配送。

2 DRL-SIA 算法

2.1 算法框架

DRL-SIA算法使用水波优化算法作为基本框架,训练完成的DQN Agent模型作为决策者,两者共同组建DRL-SIA框架模型,如图2所示。DRL-SIA框架算法对待同一类问题训练达到一定次数时,可作为解决此类组合优化问题的一般方法,具体的算法过程如下所示。

步骤1:读取环境信息完成问题模型初始化,初始化DRL-SIA算法波长 λ 、波高 h 、碎浪系数 a 、最大迭代次数 t_{max} 。

步骤2:进行种群初始化、计算每个解 x 的适应度值 $f(x)$,找出其中的最优解 x^* 。

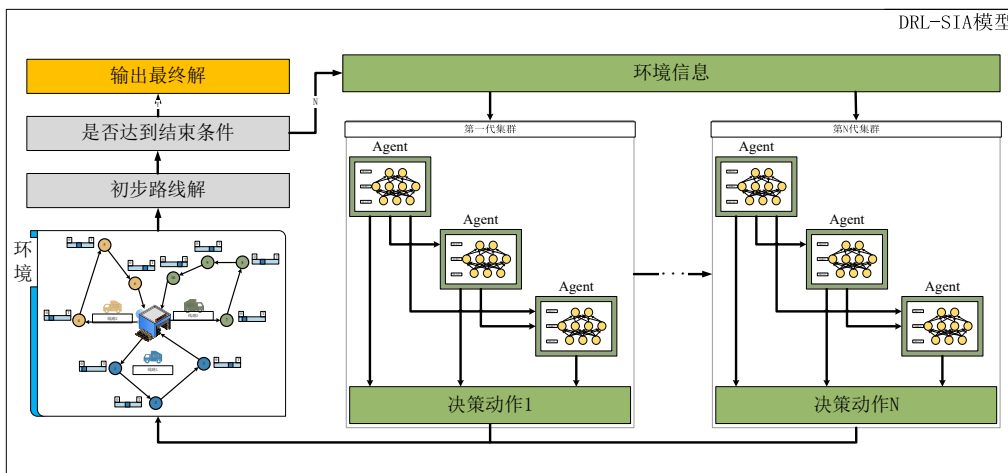


Fig. 2 DRL-SIA framework model

图 2 DRL-SIA 框架模型

步骤 3: 设迭代次数 $t = 0$ 。

步骤 4: 假如 $t = t_{max}$, 算法结束, 返回最优解 x^* 。

步骤 5: 对种群中每个解 x 使用 DQNAgent 执行传播操作。

步骤 6: 如果生成的 x' 满足 $f(x') > f(x)$, 则使用 x' 替换 x ; 如果 $f(x') > f(x^*)$, 则用 x' 替换 x^* , 并对 x' 使用 DQNAgent 执行碎浪操作。

步骤 7: 如果 $h \neq 0$, 将 x 的 h 值减 1; 如果 $h = 0$, 对 x 使用 DQNAgent 执行折射操作并重置 h 值。

步骤 8: 更新所有解波长, $t = t + 1$ 并跳转到步骤 4。

2.2 深度强化学习

深度强化学习更像是人类学习新事物的模式, 针对目标任务和环境, 持续从错误中学习并改进策略完成指定任务。经典的深度强化学习包含两个重要组成模块, 即环境和智能体。其中, 智能体就是决策者, 以环境状态为输入预测最佳动作, 以此改变环境并获得环境奖励。

多约束车辆路径问题是 NP 难问题, 由于收货点位置、数量、收货时间不断变化, 问题求解更复杂^[16]。目前, 深度学习是人工智能研究的一个热门领域, 是数据驱动的机器学习方法, 但根据已有历史数据推测将来某一事件发生的概率, 不太适用于农资配送中动态的车辆路径调度。强化学习是不断连续变化的过程, 根据当前时刻与上一时刻的状态和动作, 推断下一时刻某一动作发生的概率, 能进行车辆路径调度决策。然而, 由于车辆路径复杂及状态连续变化过快, 离散化后状态空间集合很大, 此时 Q-Learning 等传统强化学习方法也难以在内存受限的情况下维护庞大的 Q 表^[17]。

深度强化学习能使用神经网络代替 Q 表及经验回放机制解决训练样本问题, 因此结合强化学习决策和深度学习感知能力, 是解决复杂感知决策问题的有效办法。离散组合优化使用群智能算法随机性和深度学习感知能力相结合, 使其具备更优的求解精度、更快的收敛速度和更强的跳出局部最优能力。DQN (Deep Q Network) 是最常

用的深度强化学习算法, 本文使用 DRL-SIA 框架解决问题, 将训练后的智能体与群智能算法结合, 使用其代替原算法进行决策选择以提升寻优速度^[18]。下面介绍 DRL-SIA 框架中 DQNAgent 的状态空间、动作空间、奖励函数的表示方法。

2.2.1 状态空间

本文目标是 minimized 农资配送成本, 考虑到将收货点划分为不同车辆会引起车辆路径变化, 从而影响农资配送成本。因此, 将车辆路径状态形式化表示为环境状态, 主要由配送中心信息、收货点信息、车辆信息 3 部分组成。

2.2.2 动作空间

路径规划分配器的任务是将每笔订单分配到合适的路径中, 假设收货点数量为 m , 动作空间 $A = \{P_1, P_2, P_3, \dots, P_i, \dots, P_m\}$, 表示所有收货点顺序码。首先, 根据约束条件对顺序码切割路径, 例如切割后第一条完整的车辆规划路径为 $[d_1, 28, 15, 21, 14, 8, d_1]$, 其含义为: 配送中心 d_1 > 收货点 28 >> 收货点 15 >> 收货点 21 >> 收货点 14 >> 收货点 8 >> 配送中心 d_1 ; 然后, 其余路线按照相同方法, 根据顺序码分配车辆进行配送。

2.2.3 奖惩函数

奖惩函数的设计在深度强化学习中是极其重要的一环, 将任务目标具体化和数值化, 引导 DQNAgent 选择动作池中的动作策略。回报函数设计是否符合目标需求将决定 DQNAgent 能否学到期望策略, 影响着算法的收敛速度和最终性能。鉴于当前阶段的优化目标为配送成本最小化, 如果经常选择动作效果较差的动作策略, 势必会造成迭代结束时无法找到最优解, 因此将奖励函数和惩罚函数定义为:

$$R = (1 + \alpha) \times R \tag{17}$$

$$R = \beta \times R \tag{18}$$

式中: R 表示策略池选择概率; α 表示 DQNAgent 学习率; β 表示 DQNAgent 折扣因子。

2.3 群智能优化算法

群智能优化算法中水波优化算法具有控制参数少、操作简单、容易实现等优点,在目标优化、工程计算、组合优化等方面取得了较大成功^[19,20]。在水波优化算法中,水波有传播操作、折射操作和碎浪操作3种形式对空间进行搜索。其中,传播操作即水波向周围扩散开来;折射操作是水波高度到达0 m时改变传播方向;碎浪操作是水波高度较高时水波破碎形成浪花。

由此可见,水波传播贯穿整个算法流程,而折射只发生在水波高度减少至0 m时,碎浪则发生在水波过高时。为此,本文根据实际问题及其子问题设计动作池的动作策略,利用训练DQNAgent模块对动作池中的动作策略进行选择,在水波优化算法中使用训练好的DQNAgent模型在3种操作中进行寻优操作。

2.3 动作池决策动作设计

动作池决策设计分为两类,随机性决策动作和目的性决策动作。其中,随机性决策动作可跳出局部最优;目的性决策动作可加快收敛速度。本文共设计10种决策动作,随机性决策动作包括全随机、随机惋惜、随机贪心和寻差随机修复大邻域搜索,目的性决策动作包括自我学习搜索策略、最优学习搜索策略、近邻域搜索策略、变邻域搜索策略和寻差惋惜与寻差贪心大邻域搜索策略。

2.4 DQNAgent训练

DQNAgent模型在初始情况下动作池进行公平选择决策动作,通过神经网络选择决策动作,动作完成后根据结果对DQNAgent模型反馈当前状态和奖惩值。DQNAgent模型如图3所示。每个回合DQNAgent的学习训练的过程如下所示。

步骤1:初始化状态空间,包括初始化配送中心信息、收货点信息、车辆信息。

步骤2:若未达到结束状态,则通过神经网络选择一个决策动作。

步骤3:使用步骤2选中的决策动作将候选解进行调度。

步骤4:计算本次调度获得的奖惩值。

步骤5:计算当前环境状态。

步骤6:将当前状态和奖惩值保存到记忆池。

步骤7:累计奖惩值,若已经完成则转步骤8,否则转步骤2。

步骤8:若记忆池中样本数量已经达到规定阈值则训练网络。

步骤9:记录当前回合奖惩值,并统计每个路线的位置信息、时间信息、成本信息。

3 实验结果与分析

3.1 实验环境

实验语言环境为Python 3.8(Anaconda3),操作系统为Windows 10,实验机的配置为3.6 GHz主频的Core i5处理器。

3.2 算例设计

本文设计了4种不同规模的测试算例,每个规模设计10种算例以验证DRL-SIA框架算法在求解DCVRPTW时的性能。收货点规模分为小规模30个收货点、中规模50个收货点、大规模100个收货点和超大规模200个收货点,配送中心的数量设置为1,收货点时间窗长度设置为60-720之间的随机数,收货点时间窗设置为0-1 440之间的随机数,收货点需求量设置为1-10之间的随机数,收货点的服务时间设置为15,配送中心的时间窗设置为0-1 440,车辆速度设置为1,车辆单位距离使用成本设置为1,车辆启

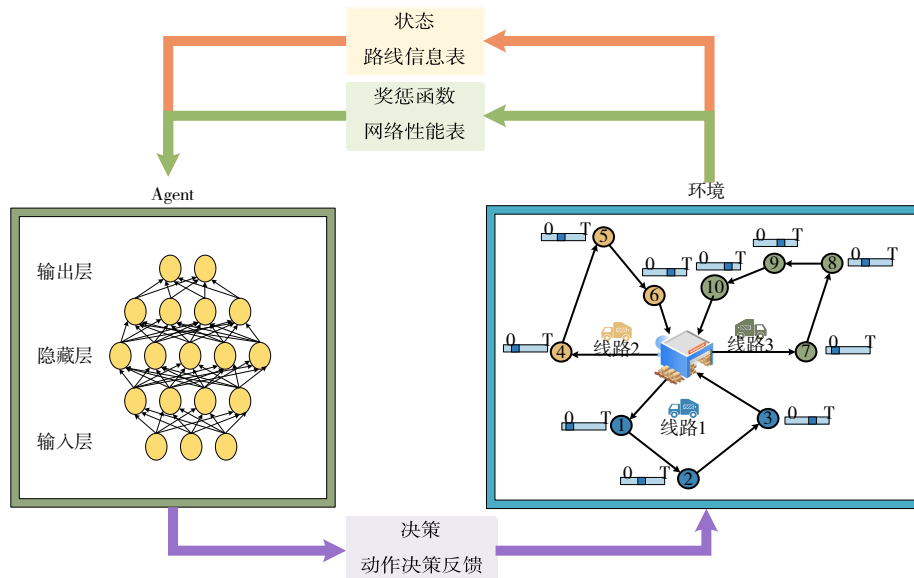


Fig. 3 DQNAgent model

图3 DQNAgent模型

动成本设置为 100, 车辆载重设置为 250, 车辆最大行驶距离设置为 400。实验终止条件设置为达到预定运行时间, 具体实验算例如表 2 所示。

Table 2 Example of different sizes
表 2 不同规模算例

算例号	收货点数量	时间窗长度	时间窗范围	需求量	服务时间	车辆速度	使用成本	固定成本	行驶距离	载重量
算例 1	30	60-720	0-1 440	1-10	15	1	1	100	400	250
算例 2	50	60-720	0-1 440	1-10	15	1	1	100	400	250
算例 3	100	60-720	0-1 440	1-10	15	1	1	100	400	250
算例 4	200	60-720	0-1 440	1-10	15	1	1	100	400	250

3.3 参数调优

深度强化学习网络模型参数包括模型参数和超参数, 模型参数通常由数据驱动调整, 例如卷积核具体核参数、神经网络的权重等, 超参数则无需数据驱动, 而在训练前或训练中人为调整, 例如学习率、折扣因子等。超参数通常先按经验设定初始值, 然后根据训练效果进行调优。

由于学习率直接控制训练网络梯度更新的量级, 因此在所有的超参数中学习率最重要, 图 4 给出了 DRL-SIA 框架算法在不同学习率(α)下的目标函数收敛曲线, 由此可见当 $\alpha=0.1$ 时目标函数求解精度最优。

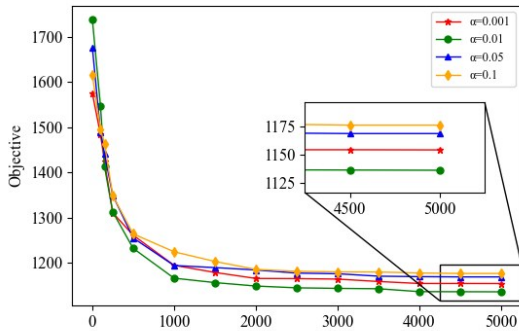


Fig. 4 Convergence curves for different learning rates
图 4 不同学习率情况下收敛曲线

此外, 折扣因子也是深度强化学习中一个重要的超参数, 图 5 给出了 DRL-SIA 框架算法在不同折扣因子(β)情况下的目标函数收敛曲线。由此可见, 当 $\beta=0.9$ 时目标函数更平稳, 收敛速度和求解精度优于其他对照组。

本文根据训练效果确定两个超参数指标, 得到一组良

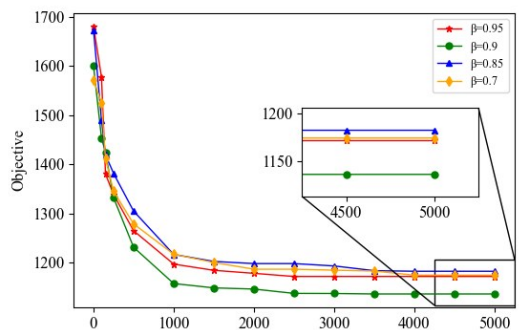


Fig. 5 Convergence curves for different discount factors
图 5 不同折扣因子情况下收敛曲线

好的超参数指标可提升深度强化学习的效率和寻优能力。经过参数调优发现, DRL-SIA 框架的重要超参数学习率设为 0.1、折扣因子设为 0.9、目标网络更新频率设为 100 时最优。

3.4 算法性能测试

比较算法选择当前效果较好的混合算法 PSO-GA、GA-VNS 和本文框架的原始算法水波优化算法。在相同时间通过算法稳定性、求解精度、收敛速度和是否具有显著差别来测试本文所提算法的求解性能和统计性能。

3.4.1 不同规模下算法稳定性

根据式(19)得到各算法在各实验算例中的 ARPD 值。

$$APRD = Average_{ins} \left(\frac{1}{R} \sum_{i=1}^R \frac{C_i - C_{opt}}{C_{opt}} \times 100\% \right) \quad (19)$$

式中: C_i 表示在第 i 次实验中得到的最优目标值; C_{opt} 表示多次运行得到的最优目标值; R 表示算法在单个示例上运行的次数; ins 代表每种示例的数量 $ins = 1, 2, 3, \dots, 10$ 。

表 3 给出了收货点数量分别为 30、50、100、200 时的 ARPD 值, 加粗数据代表最优数据。实验表明, DRL-SIA 框架算法稳定性明显优于其他算法, 对于求解 DCVRPTW 问题的稳定性良好, 可更好地跳出局部最优, 不会因某个局部最优而导致本次求解与其他结果差距过大。

Table 3 Results of ARPD of different methods at different scales

规模	DRL-SIA	GA-VNS	PSO-GA	WWO
30	0.00	10.10	2.74	9.75
50	2.85	18.51	6.17	10.01
100	6.07	6.18	6.20	9.44
200	1.82	2.10	3.92	6.32

3.4.2 不同规模下求解精度

DRL-SIA 框架在不同规模的 DCVRPTW 测试算例中稳定性和求解精度结果如图 6 所示。实验表明, 所提模型的稳定性和求解精度均优于其他算法, 原因为 DRL-SIA 框架的 DQNAgent 策略可引导算法搜索过程, 根据问题特性进行探索以提升算法求解能力。

3.4.3 不同规模下收敛速度

本文为了验证 DRL-SIA 框架的收敛速度, 对 4 种算法在不同规模 (30、50、100、200) DCVRPTW 测试算例中的实验结果进行统计分析, 如图 7 所示。由此可见, DRL-SIA 框架在收敛速度方面相较于其他算法优势更大, 原因为本文算法通过问题特性选择搜索策略, 在收敛速度上领先其他算法。

3.4.4 不同规模下算法显著性

为了进一步测试 DRL-SIA 框架的统计性能, 分别在收货点数量为 30、50、100、200 的情况下得到 Wilcoxon 秩和检验结果, 如表 4—表 7 所示。由此可知, DRL-SIA 框架相较于其他算法均具有显著性差异。

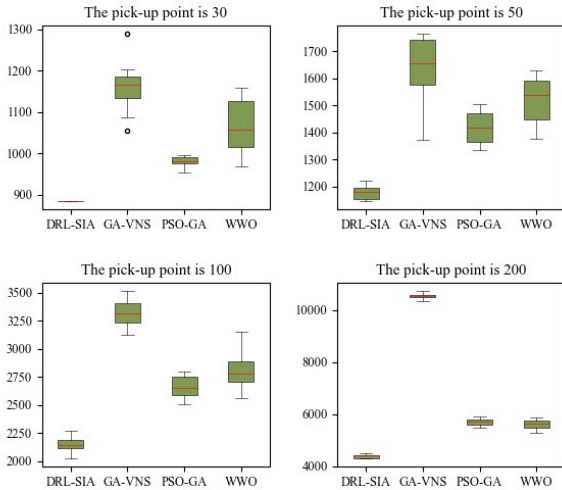


Fig. 6 Solution accuracy of each method at different scales
图6 不同规模下各方法求解精度

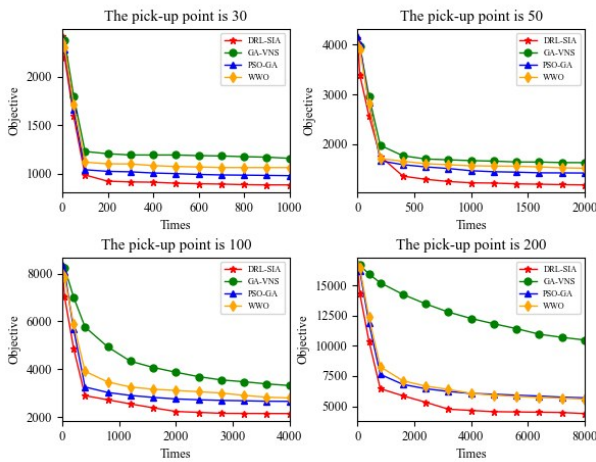


Fig. 7 Convergence curves of various methods at different scales
图7 不同规模下各方法收敛曲线

Table 4 Results of Wilcoxon rank sum test when the number of receiving points is 30
表4 收货点数量为30情况下Wilcoxon秩和检验结果

DRL-SIA vs	R+	R-	Z	p Value	$\alpha = 0.05$	$\alpha = 0.1$
GA-VNS	10	0	-2.80 306	5.06E-03	yes	yes
PSO-GA	10	0	-2.80 306	5.06E-03	yes	yes
WWO	10	0	-2.80 306	5.06E-03	yes	yes

Table 5 Results of Wilcoxon rank sum test when the number of receiving points is 50
表5 收货点数量为50情况下Wilcoxon秩和检验结果

DRL-SIA vs	R+	R-	Z	p Value	$\alpha = 0.05$	$\alpha = 0.1$
GA-VNS	10	0	-2.80 306	5.06E-03	yes	yes
PSO-GA	10	0	-2.80 306	5.06E-03	yes	yes
WWO	10	0	-2.80 306	5.06E-03	yes	yes

3.5 模型计算结果

3.5.1 不同规模下经济成本

从经济成本角度比较4种算法结果,结果如图8、表8所示。由此可知,DRL-SIA框架在经济成本中全面优于比较算法,能减少农用物资电子商务平台的物流支出。

Table 6 Results of Wilcoxon rank sum test when the number of receiving points is 100
表6 收货点数量为100情况下Wilcoxon秩和检验结果

DRL-SIA vs	R+	R-	Z	p Value	$\alpha = 0.05$	$\alpha = 0.1$
GA-VNS	10	0	-2.80 306	5.06E-03	yes	yes
PSO-GA	10	0	-2.80 306	5.06E-03	yes	yes
WWO	10	0	-2.80 306	5.06E-03	yes	yes

Table 7 Results of Wilcoxon rank sum test when the number of receiving points is 200
表7 收货点数量为200情况下Wilcoxon秩和检验结果

DRL-SIA vs	R+	R-	Z	p Value	$\alpha = 0.05$	$\alpha = 0.1$
GA-VNS	10	0	-2.80 306	5.06E-03	yes	yes
PSO-GA	10	0	-2.80 306	5.06E-03	yes	yes
WWO	10	0	-2.80 306	5.06E-03	yes	yes

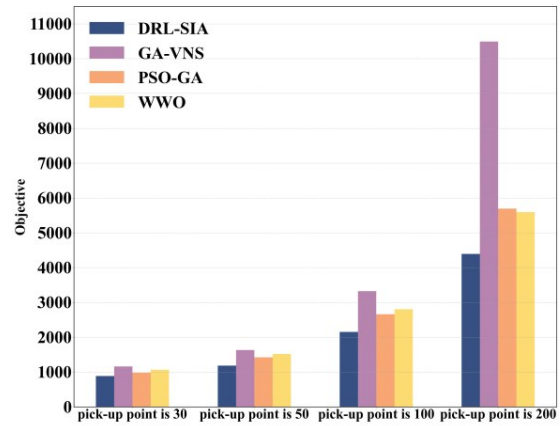


Fig. 8 Economic costs of various methods at different scales
图8 不同规模下各方法经济成本

Table 8 Economic costs of various methods at different scales
表8 不同规模下各方法经济成本 (%)

DRL-SIA vs	30pints	50pints	100pints	200pints
GA-VNS	23.69	27.49	35.29	58.18
PSO-GA	9.75	16.80	19.22	22.95
WWO	16.84	22.12	23.36	21.53

3.5.2 不同规模下行驶距离

行驶距离是物流供应链体系中的重要指标,实验结果如图9、表9所示。由此可知,除了在小规模的问题上与PSO-GA算法基本持平之外,其余结果均优于其他算法。

3.5.3 不同规模下行驶时间

行驶时间也是物流供应链体系中的重要指标,实验结果如图10、表10所示。由此可知,除了在小规模问题上略高于PSO-GA算法之外,其余结果均优于其他算法。

3.5.4 不同规模下车辆数量

由于投入到供应链中的车辆数量代表农用物资电子商务平台投入固定资产,投入越多风险越高,各算法结果如图11、表11所示。由此可知,本文框架在车辆数量中全面优于比较算法,能减少农用物资电子商务平台的固定资产投入,降低金融风险。

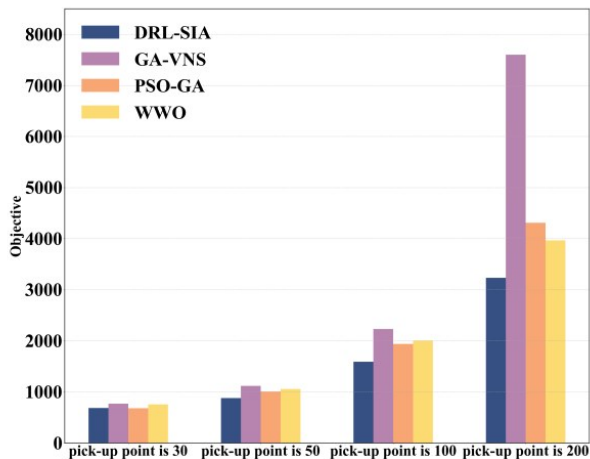


Fig. 9 Driving distance of various methods at different scales
图 9 不同规模下各方法行驶距离

Table 9 Driving distance of various methods at different scales
表 9 不同规模下各方法行驶距离 (%)

DRL-SIA vs	30pints	50pints	100pints	200pints
GA-VNS	11.15	21.26	28.78	57.45
PSO-GA	-0.65	12.72	18.10	25.00
WWO	9.18	16.61	20.71	18.43

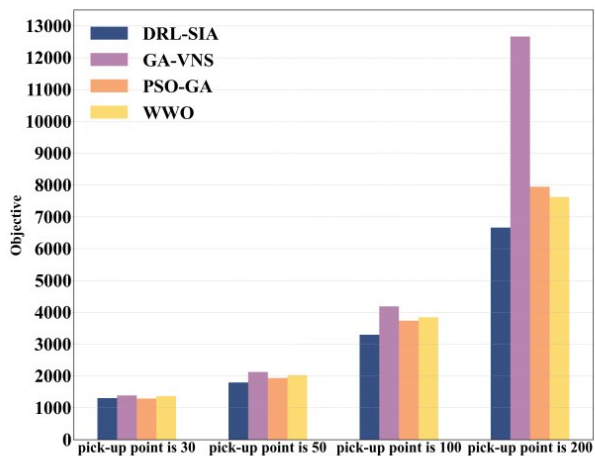


Fig. 10 Travel time of various methods at different scales
图 10 不同规模下各方法的行驶时间

Table 10 Travel time of various methods at different scales
表 10 不同规模下各方法的行驶时间 (%)

DRL-SIA vs	30pints	50pints	100pints	200pints
GA-VNS	5.94	15.55	21.25	47.38
PSO-GA	-1.27	7.21	11.81	16.16
WWO	4.46	11.57	14.17	12.61

4 结语

本文提出一种新的 DRL 模型求解 VRP, 该模型由生成农资订单配送解决方案的参与者和训练参与者组成。首先, 通过改进奖惩函数优化参与者参数减少训练方差, 然

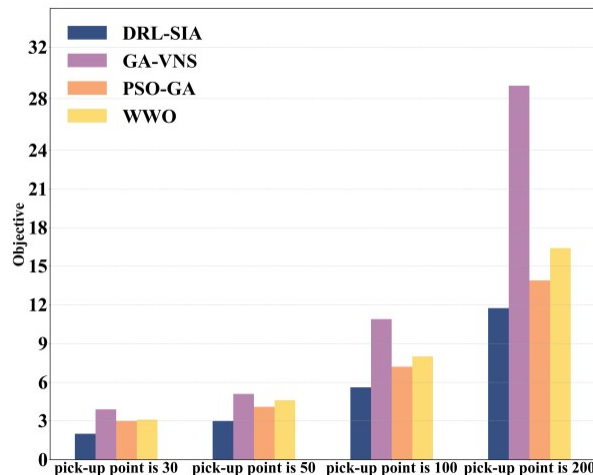


Fig. 11 Number of vehicles using different methods at different scales
图 11 不同规模下各方法车辆数量

Table 11 Number of vehicles using different methods at different scales

DRL-SIA vs	30pints	50pints	100pints	200pints
GA-VNS	48.72	41.18	48.62	59.48
PSO-GA	33.33	26.83	22.22	15.47
WWO	35.48	34.78	30.00	28.35

后设计了一个农资配送模拟器帮助 DRL 模型训练, 以此提出一种由 DRL 和群智能算法组成的 DRL-SIA 框架求解 DCVRPTW。

实验表明, DRL-SIA 框架相较于比较算法求解精度更高、算法稳定性更强、收敛速度更快, 可解决传统算法无法解决的大规模问题。由于算法框架集成了 10 种优化策略, 可应对各种陷入局部最优的情况。同时, 基于深度强化学习训练智能作出更适合问题特性的决策, 使该框架可选择更出色的路由策略, 可将结合 DRL 与启发式方法作为解决组合优化问题的一般方法。

未来, 一方面利用先进的强化学习技术, 例多智能体协作的注意力机制提升 DRL 模型性能; 另一方面将 DRL-SIA 框架应用于更复杂的物流路由问题和其他组合优化问题。

参考文献:

[1] Office of the Leading Group of the Third National Land Survey of the State Council. Third national land survey main data bulletin[N]. People's Daily, 2021-08-27(17).
国务院第三次全国国土调查领导小组办公室. 第三次全国国土调查主要数据公报[N]. 人民日报, 2021-08-27(17).

[2] Bureau of Statistics of the People's Republic of China. China statistical yearbook[M]. Beijing: China Statistics Press, 2022.
中华人民共和国统计局. 中国统计年鉴[M]. 北京: 中国统计出版社, 2022.

[3] WANG X, GAO B, AN K. Research on guizhou agricultural material logistics distribution mode from the perspective of new retail[J]. Logistics

- Sci-tech, 2020, 43(1): 45-47.
汪旭, 高波, 安旭. 新零售视角下贵州农资物流配送模式研究[J]. 物流科技, 2020, 43(1): 45-47.
- [4] SHI G L, SHEN X Y, GU L C, et al. Research progress on chipless RF cross-domain sensing for smart agriculture[J]. Journal of Agricultural Engineering, 2023, 39(7): 10-23.
时国龙, 沈心怡, 辜丽川, 等. 面向智慧农业的无芯片射频跨域感知研究进展[J]. 农业工程学报, 2023, 39(7): 10-23.
- [5] FENG Y, FENG Y C, ZHANG J X, et al. Decision-making for low-carbon supply chain operation of fresh food considering carbon emissions from production and transportation[J]. Journal of Systems Management, 2024, 33(1): 1-15.
冯颖, 冯仰超, 张景雄, 等. 考虑产运环节碳排放的生鲜低碳供应链运营决策[J]. 系统管理学报, 2024, 33(1): 1-15.
- [6] DOMINGUEZ-MARTIN B, RODRÍGUEZ-MARTÍN I, SALAZAR-GONZÁLEZ J. An efficient multistart heuristic for the driver and vehicle routing problem [J]. Computers & Operations Research, 2023, 150: 106076.
- [7] LI J H. Research on optimization method of urban logistics and distribution vehicle path [D]. Changchun: Changchun University of Science and Technology, 2021.
李佳慧. 城市物流配送车辆路径优化方法研究[D]. 长春: 长春理工大学, 2021.
- [8] BORCINOVA Z. Kernel search for the capacitated vehicle routing problem [J]. Applied Sciences-Basel, 2022, 12(22): 11421.
- [9] MAHMOOD N. Solving capacitated vehicle routing problem using meerkat clan algorithm[J]. International Arab Journal of Information Technology, 2022, 19(4): 689-694.
- [10] BIRECIK Z A, ÖZGEN D. An interactive possibilistic programming approach for green capacitated vehicle routing problem[J]. Neural Computing & Applications, 2023, 35: 9253-9265.
- [11] CAI J S, WANG P, SU S Q. A dynamic space reduction ant colony optimization for capacitated vehicle routing problem [J]. Soft Computing, 2022, 26(17): 8745-8756.
- [12] DALBAH L M, AL-BETAR M A, AWADALLAH M A, et al. A modified coronavirus herd immunity optimizer for capacitated vehicle routing problem[J]. Journal of King Saud University-Computer and Information Sciences, 2022, 34(8): 4782-4795.
- [13] RATTANAMANEE T, NANTHAVANIJ S. Heuristic procedure for bi-capacitated multiple-trip vehicle routing problem [J]. European Journal of Industrial Engineering, 2022, 16(3): 294-316.
- [14] QIN H B, LI C X, TANG H T, et al. Research on MOMA-based reentrant hybrid flow shop scheduling problem [J]. Journal of System Simulation, 2024, 36(1): 131-148.
秦红斌, 李晨晓, 唐红涛, 等. 基于MOMA的可重入混合流水车间调度问题研究 [J]. 系统仿真学报, 2024, 36(1): 131-148.
- [15] BAO D Q, GE N, ZHAI S M, et al. Boolean satisfiability prediction based on message-passing relational networks [J]. Journal of Software, 2022, 33(8): 2839-2850.
包冬庆, 葛宁, 翟树茂, 等. 基于消息传递关系网络的布尔可满足性预测[J]. 软件学报, 2022, 33(8): 2839-2850.
- [16] LIN J, YE J X, LIU W W, et al. A multimodal differential evolutionary algorithm for solving vehicle path problems with capacity constraints [J]. Computer Applications, 2023, 43(7): 2248-2254.
林剑, 叶璟轩, 刘雯雯, 等. 求解带容量约束车辆路径问题的多模态差分进化算法[J]. 计算机应用, 2023, 43(7): 2248-2254.
- [17] LI W, ZHANG X D, JIANG X F, et al. Research on robot path planning based on improved reinforcement learning [J]. Manufacturing Automation, 2023, 45(3): 148-151, 172.
李威, 张晓东, 姜学峰, 等. 基于改进强化学习的机器人路径规划研究[J]. 制造业自动化, 2023, 45(3): 148-151, 172.
- [18] XIE Y, ZHENG S J, CHENG N S, et al. Robot path planning research based on improved reinforcement learning [J]. Manufacturing Automation, 2023, 45(3): 148-151, 172.
谢勇, 郑绥君, 程念胜, 等. 基于改进DQN算法的路径规划方法研究 [J]. 制造业自动化, Manufacturing Automation, 2023, 45(3): 148-151, 172.
- [19] ZHENG Y J. Water wave optimization: a new nature-inspired meta-heuristic [J]. Computers & Operations Research, 2015, 55: 1-11.
- [20] ZHANG J F. Application of water wave optimization algorithm in allocation scheduling problem [D]. Hangzhou: Zhejiang University of Technology, 2017.
张杰峰. 水波优化算法在分配调度问题中的应用[D]. 杭州: 浙江工业大学, 2017.

(责任编辑: 刘嘉文)