

基于强化学习的带软时间窗多行程 绿色车辆路径优化研究

姚利军,王可君[†]

(湖南省长株潭烟草物流有限责任公司,湖南 长沙 410004)

摘 要:为了助力物流行业响应“碳达峰”和“碳中和”建设目标,提速绿色物流产业的建立与发展,首先综合考虑油耗、碳排放、人力、车辆、用户体验等因素,构建带软时间窗约束的多行程绿色车辆路径优化模型。然后综合考虑 PinSAGE 图网络、TRPO 和 GAE 方法来改进 Actor-Critic 的深度强化学习优化算法,最后采用 Actor-Critic 算法对模型对绿色多行程车辆路径方案求解。实验表明,提出的求解方法能高效规划绿色车辆路径,进而显著减少物流成本,实现物流企业经济效益与环境效益的双重优化。

关键词:绿色物流;软时间窗;深度强化学习;Actor-Critic 框架

中图分类号:F252;U121

文献标识码:A

Reinforcement Learning Based Path Optimization for Multi-Trip Green Vehicles with Soft Time Window

YAO Lijun, WANG Kejun[†]

(Changzhutan Tobacco Logistics Co., Ltd. of Hunan, Changsha, Hunan 410004, China)

Abstract: In order to assist the logistics industry in achieving its goal of peak carbon dioxide emissions and carbon neutrality, construction must be facilitated, and a green logistics industry must be rapidly established and developed. Firstly, a multi-trip green vehicle path optimization model with soft time window constraints is constructed by comprehensively considering the factors of fuel consumption, carbon emission, manpower, vehicles, and user experience. Subsequently, the PinSAGE graph network, TRPO, and GAE methods are considered collectively to enhance the deep reinforcement learning optimization algorithm of Actor-Critic. Ultimately, the Actor-Critic algorithm is employed to address the model for the green multi-trip vehicle path scheme. Experimental evidence indicates that the solution method proposed in the paper is an effective means of planning green vehicle routes, which in turn has the potential to significantly reduce logistics costs and realise the dual optimisation of economic and environmental benefits for logistics enterprises.

Key words: green logistics; soft time window; deep reinforcement learning; Actor-Critic framework

电子商务的兴盛推动了物流行业发生剧变,亟须向精细化、智能化、绿色转型,以解决成本控制与效率提升的紧迫问题。围绕物流行业降本增效与绿色环保理念,Bruglieri 等^[1]为高效求解绿色车辆路径问题,通过生成和组合可行的非支配路径来最

小化总行驶距离,同时考虑了车辆在行驶过程中需要在替代燃料站停留加油的限制。张静怡等^[2]以车辆固定、运输相关、未满足需求的惩罚、碳排放等成本加权和为优化目标,建立了基于低碳排放的应急物流车辆路径优化模型,解决灾害情况下货物的

运输。张金良等^[3]考虑碳排放背景下针对客户点增减、客户需求量变化及配送路线交通中断三种优化问题,动态优化配送车辆路径以获得最低配送成本。利用软时间窗口约束来平衡成本和用户体验,刘虹等^[4]为解决配送过程中客户需求动态性和不确定性问题,引入随机机会约束规划模型,并提出了“实时柔性点”的路径调整策略,能够适应客户取送货需求的随机性并进行实时优化调整。范厚明等^[5]针对模糊需求和时间窗口的车辆路径优化问题,改进交叉算子操作优化混合遗传算法,同时引入擂台法则提高种群多样性,促进求解效率的提高。Tas 等^[6]考虑软时间窗口和行驶时间不确定性的情况,利用禁忌搜索方法快速生成配送路径,同时兼顾了运输和服务成本。He 等^[7]通过改进蚁群算法的信息素更新策略来弥补算法初期信息素缺乏的问题,并设计了两个变邻域搜索算子以避免算法陷入局部最优解,来平衡配送过程中的车辆服务灵活性。为了加快物流效率与绿色物流产业发展,构建了许多考虑碳排放、多行程以及用户体验的软时间窗口约束等车辆路径模型。周鲜成等^[8]考虑时间窗口依赖与多行程的货物配送问题,考虑车辆行驶速度、实时载重等因素建立了油耗和碳排放测度模型;根据配送时间和顾客满意度之间的关系来优化车辆配送方案。Eskandarpour 等^[9]探讨容量和行驶范围对求解的影响,并且更贴合实际地考虑了碳排放约束,建立了最小化总成本和最少化二氧化碳排放量的双优化目标的整数规划模型并采用改进的大邻域搜索算法进行求解。深度强化学习方法,作为一种高效的车辆路径求解策略,可为物流行业提供更加智能、高效的解决方案。周雅兰等^[10]提出了 Memetic 深度强化学习算法求解带时间窗口约束的同时取送货车辆路径问题,在充分利用已有人工启发式算法的基础上进一步优化求解效果。罗佳等^[11]提出基于残差图卷积和多头注意力融合的深度强化学习算法构建可行解序列,提高需求可拆分车辆路径问题的求解效率。Xin 等^[12]采用多解码器注意力模型解决车辆路径优化问题,通过训练多个不同的策略有效提高发现优秀解的概率。Raajan 等^[13]将车辆及其所处环境的即时状况直接转化为驾驶操作,采用深度确定性策略梯度方法来优化车辆物流配送流程。

综上所述,现有研究存在以下不足:第一,绿色车辆路径优化问题多集中于多车场与多车型,存在车辆调度冗余、资源浪费、成本消耗大的问题;第二,模型求解方法大多采用启发式算法,致使求解难度较大、求解效率低等问题;第三,较少利用强化学习算法求解,且求解中大多与环境交互不彻底,

优化策略不会随求解效果自适应调整等。针对上述不足,本文在综合考虑油耗、碳排放、人力、车辆、用户体验等因素,构建了带软时间窗口多行程绿色车辆路径模型 (Green Multi-Trip Vehicle Routing Problem with Time Windows, GMTVRPTW), 利用基于 Actor-Critic 的深度强化学习算法求解,实现在最小碳排放成本条件下的物流总成本最小化的目标。

1 问题描述和模型构建

1.1 问题描述

GMTVRPTW 问题中采用单一车型建立配送编队,按照运输任务规划和时间窗口限制完成多个客户点的配送任务并最终返回物流中心。在优化车辆时,已知物流中心与各个配送点的具体位置、各点间距离、各点所需要商品数量、服务时间窗口等信息,在满足客户需求与配送限制的情况下,实现兼顾送货总成本和碳排放的车辆路径规划。为方便研究,作出如下几点假设:(1)只有一个物流中心;(2)物流中心拥有所有客户需要的货物;(3)已知各点时间窗口;(4)已知物流中心和每个客户点的地理坐标;(5)已知每个客户点的需求量;(6)物流中心拥有足够多的单一型号车辆;(7)司机完全遵从系统设计的路线;(8)车辆从物流中心出发完成配送任务后直接返回中心;(9)车辆从物流中心出发时所载货物总量不超载。

1.2 符号和参数定义

本文构建的带软时间窗口多行程绿色物流车辆路径模型中符号和参数定义如下:

(1)绿色低碳需求下车辆碳排放量和油耗量的计算

对于碳排放和油耗,本文参考 Hickman^[14]提出的 MEET 模型进行碳排放计算方式,其碳排放率计算函数如公式(1)所示,考虑车辆行驶速度和载重量对碳排放率的影响,可用公式(2)~(4)计算碳排放率函数(kg/km)与油耗率(L/kg)之间的关系:

$$\epsilon(v) = \omega_0 + \omega_1 v + \omega_2 v^2 + \omega_3 v^3 + \frac{\omega_4}{v} + \frac{\omega_5}{v^2} + \frac{\omega_6}{v^3} \quad (1)$$

$$LC = \chi_0 + \chi_1 \gamma + \chi_2 \gamma^2 + \chi_3 \gamma^3 + \chi_4 v + \chi_5 v^2 + \chi_6 v^3 + \frac{\chi_7}{v} \quad (2)$$

$$c_{ij} = \epsilon(v)LC/1000 \quad (3)$$

$$f_{ij} = \sigma c_{ij} \quad (4)$$

其中 v 是车辆行驶速度; $\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6$ 为超参数。参考文献[15],公式(1)中的排放率最

小时的车辆行驶速度 $v^* = 71 \text{ km/h}$ 。 γ 为车辆实际载重率; $\chi_0、\chi_1、\chi_2、\chi_3、\chi_4、\chi_5、\chi_6、\chi_7$ 为超参数, σ

为车辆单位碳排放量,参考文献[16]将 ρ 值取为 0.3703 L/kg 。

表 1 符号和参数定义

符号	参数含义	符号	参数含义
N	物流中心和客户点的节点集合, $N = \{i \mid i = 1, 2, 3, \dots, n\}$, 其中 $i = 1$ 表示物流中心, 其余节点表示客户点	x_{ij}	决策变量; $x_{ij} = \begin{cases} 1 \\ 0 \end{cases}$, 其中 $x_{ij} = 1$ 表示车辆通过了点 i 到点 j 之间的路径, $x_{ij} = 0$ 则表示车辆没有通过点 i 到点 j 之间的路径
m	车辆数	T_i	车辆到达目标客户点 i 的时间
d_{ij}	点 i 与点 j 之间的距离	t_{ij}	车辆从点 i 行驶到点 j 所需的时间
l_i	客户点 i 的货物需求量	T_0	车辆从物流中心发车时刻
C	车辆最大容量	ET_i	客户点 i 最早送货时刻
PC	单位时间惩罚成本	LT_i	客户点 i 最晚送货时刻
v	车辆行驶速度	α	车辆的单位时间使用成本
f_{ij}	车辆在路径 d_{ij} 中的油耗率	φ	驾驶员单位时间工资成本
c_{ij}	车辆在路径 d_{ij} 中的碳排放率	θ	车辆的发车固定成本
ρ_f	单位油耗费用	F	总物流费用
ρ_e	单位碳排放费用		

(2)带软时间窗多行程绿色车辆 GMTVRPTW 模型构建

本文以油耗与碳排放、车辆使用成本、驾驶员雇佣成本、用户体验(软时间约束)等加权和为优化目标,采用如下的 GMTVRPTW 模型:

$$F = \sum_{i=1}^n \sum_{j=1}^n x_{ij} v t_{ij} (f_{ij} \rho_f + c_{ij} \rho_e) + PC \left[\sum_{i=1}^n \max(E T_i - T_i, 0) + \sum_{i=1}^n \max(T_i - L T_i, 0) \right] + \sum_{i=1}^n \sum_{j=1}^n x_{ij} t_{ij} (\alpha + \varphi) + \sum_{i=1}^n \sum_{j=1}^n x_{ij} \theta \tag{5}$$

s. t.

$$\sum_{j=1}^n x_{1j} = \sum_{i=1}^n x_{i1} \leq 1, i, j \in 1, 2, 3, \dots, n \tag{6}$$

$$\sum_{\substack{i=1 \\ i \neq j}}^n x_{ij} = 1, j \in 2, 3, 4, \dots, n \tag{7}$$

$$\sum_{\substack{i=1 \\ j \neq i}}^n x_{ij} = 1, i \in 2, 3, 4, \dots, n \tag{8}$$

$$\sum_{i=1}^n \sum_{j=1}^n l_i x_{ij}^k \leq C_k, i, j \in 1, 2, 3, \dots, n \tag{9}$$

$$\sum_{\substack{i=1 \\ i \neq p}}^n x_{ip} = \sum_{\substack{j=1 \\ j \neq p}}^n x_{pj}, i, j \in 1, 2, 3, \dots, n, p \in 2, 3, 4, \dots, n \tag{10}$$

$$T_j = \sum_{\substack{i=1 \\ i \neq j}}^n \sum_{k=1}^m x_{ij}^k (T_i + t_{ij}), j \in 1, 2, 3, 4, \dots, n, \text{ 且 } T_1 = 0 \tag{11}$$

$$x_{ij} = \begin{cases} 1 \\ 0 \end{cases}, x_{ij} \text{ 是决策变量} \tag{12}$$

式(5)中综合考虑了碳排放、油耗成本、车辆使用与司机人力雇佣等因素,其中第一项为车辆的碳排放和油耗成本;第二项为惩罚成本,表示车辆没有被客户要求的时间段内到达,影响了客户体验感而引入的惩罚,以此平衡配送效率与客户需求的灵活性;第三项为车辆与司机的雇佣时间成本;第四项为总发车数的固定成本。式(6)表示车辆从物流中心出发后完成任务后仍需返回中心;式(7)和式(8)表示每个客户只能被一辆车服务;式(9)表示每辆车的实际载重不能超过其容量;式(10)表示车辆每到一个客户点必须卸下货物并离开;式(11)表示车辆行驶到客户点所需的时间,车辆 0 时刻从物流中心出发;式(12)表示决策变量的值域为 0 和 1。

2 改进基于 Actor-Critic 深度强化学习算法

对 GMTVRPTW 模型,利用基于 Actor-Critic 深度强化学习算法进行求解,可以处理目标函数中的复杂约束,通过其强大的学习能力和对复杂环境的适应性,为 GMTVRPTW 问题寻找高效、灵活且能够实时调整的配送策略。

2.1 车辆与环境的交互——Actor 网络设计流程

Actor 网络通过实时整合位置数据、优化路径选择、软时间窗约束、车辆容量等各项约束,利用 PinSAGE 图神经网络来捕捉客户位置、空间关系和距离等特征信息,并输入 PPO 容器中,具体过程

如下:

1) 挖掘客户节点 V_i 的位置坐标、货物需求量、服务时间等信息,并描述为特征向量 $\mathbf{x}_i^l$ 。

2) 参照各客户节点 V_i 的一阶节点和二阶节点信息,根据公式(13)更新其周围环境即邻居特征 h_i^l 。

3) 结合自身特征和邻居特征,根据公式(14)更新节点特征 $\mathbf{x}_i^{l+1}$ 。

4) 将更新后的节点特征 $\mathbf{x}_i^{l+1}$ 输入全连接网络,获得其预测状态特征值。

5) 最后,使用 LSTM 刻画访问历史捕获路径中节点顺序的相关性和依赖关系,并计算每一个客户节点 i 的访问概率,从而预测下一个最佳访问节点。

$$h_i^l = \text{AvePoo}(\{\mathbf{x}_j^l \mid V_j \in N(V_i)\}) \quad (13)$$

$$\mathbf{x}_i^{l+1} = \text{ReLU}(\mathbf{x}_i^l, h_i^l) \quad (14)$$

2.2 策略动作评估——Critic 网络设计

利用 TRPO 和 GAE 方法来处理 PPO 容器中的环境状态数据,提取优势值与奖励值,激励执行策略改进并促进指引策略调整优化方向;具体过程如下:

1) 利用 TRPO 方法从当前时间步到策略终止或 episode 结束时积累的未来奖励,根据公式(15)计算每一步的奖励值 G_t 。

2) 根据公式(16),当前状态收益与下一个状态的价值,采用单步时序差分方法 TD 估计每个状态的预测收益。

3) 根据公式(17)估计反向累积 TD 差分,得到当前时间步 t 的优势值 A_t^{GAE} 。

4) 根据公式(18)–(19),利用 TRPO 方法估算最大化期望回报以确保策略的稳定改进。

$$G_t = \sum_{k=0}^{T-t-1} \gamma^k r_{t+k+1} \quad (15)$$

$$\delta_{t+1}^v = r_{t+1} + \gamma V(S_{t+1}) - V(S_t) \quad (16)$$

$$A_t^{\text{GAE}(\gamma, \lambda)} = \sum_{l=0}^{T-t-1} (\gamma \lambda)^l \delta_{t+l} \quad (17)$$

$$r(\theta) = \frac{\pi_{\theta}(a_t \mid s_t)}{\pi_{\theta_{\text{old}}}(a_t \mid s_t)} \quad (18)$$

$$\max L^{\text{TRPO}}(\theta) = E_{a_t \sim \pi_{\text{old}}, s_t \sim \rho_{\pi_{\text{old}}}} [r(\theta) A_t^{\text{GAE}(\gamma, \lambda)}] \quad (19)$$

其中, T 是序列终点, γ 是折扣因子, r_{t+k+1} 是在 $t+k+1$ 时间步获得的奖励, δ_{t+1}^v 是 TD 误差, λ 是平滑参数控制 TD 误差和多步回报估计的权重,

一般取值范围在 0 和 1 之间。

2.3 深度强化学习算法训练模型

利用梯度下降法更新神经网络参数,具体过程如下:

1) 利用 PPO 算法,通过 clip 技术控制策略改变的幅度从而保障学习的稳定性和高效性;

2) 根据每一步决策的优势值和奖励值,权衡策略更新的平滑性和价值函数精度,引入熵损失激励探索策略的多样性。PPO 算法设计中利用 Clip 技术的策略损失、价值损失、熵损失和联合损失如公式(20)–(23)所示:

$$L^{\text{CIP}}(\theta) = E_t[\min(r(\theta) A_t, \text{clip}(r(\theta), 1 - \epsilon, 1 + \epsilon) A_t)] \quad (20)$$

$$L^{\text{VF}}(\theta) = E_t[(V_{\theta}(s_t) - G_t)^2] \quad (21)$$

$$L^{\text{E}}(\theta, \theta) = -\beta E_t[\log \pi_{\theta}(a_t \mid s_t)] \quad (22)$$

$$L(\theta, \theta) = L^{\text{CIP}}(\theta) + C_1 L^{\text{VF}}(\theta) - C_2 L^{\text{E}}(\theta, \theta) \quad (23)$$

A_t 是归一化的优势估计, θ 是 clip 技术的阈值, $\text{clip}(r(\theta), 1 - \epsilon, 1 + \epsilon)$ 是一个裁剪函数,用于限制策略的更新幅度。 E_t 是一个经验期望, β 、 C_1 和 C_2 分别是熵系数和平衡损失的系数。

2.4 基于深度强化学习算法求解 GMTVRPTW 整体流程

本文采用的基于 Actor-Critic 深度强化学习算法求解带软时间窗多行程绿色车辆路径模型 GMTVRPTW 工作流程,具体如图 1 所示:

Actor 工作流程如下:

1) 提取距离、时间窗余量、约束信息等实时状态信息,增强信息适应性;

2) 然后将状态信息送入由 PinSAGE 图网络构成的 Actor 网络中,估计每个客户访问概率和下一步可采取的动作集;

3) 再根据概率分布随机选取一个动作,更新状态、奖励等信息;

4) 重置环境,继续循环步骤 1)–4),直至获得合适的解;

5) 将求得的解的相关信息数据存入 PPO 算法容器。

Critic 工作流程如下:

1) 将 PPO 算法容器内数据送入 Critic 网络中;

2) 利用 TRPO 和 GAE 算法估算每一步的收益值和奖励值,具体过程参考 3.2 小节;

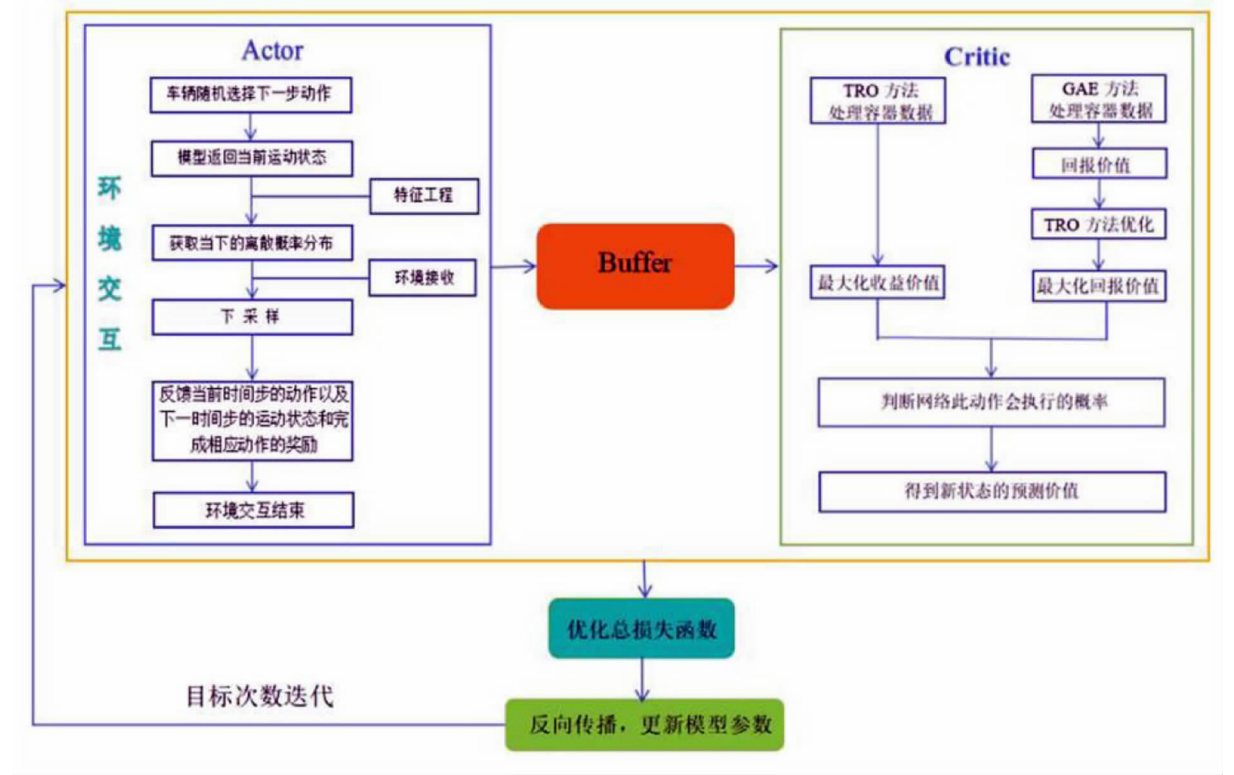


图 1 深度强化学习算法流程

3)根据当前策略与历史动作概率,通过对比分析以优化训练策略;

优化总损失函数流程如下:

1)参照公式(23)–(26),迭代更新计策略损失 $L^{CLIP}(\theta)$ 、价值损失 $L^{VF}(\phi)$ 、熵损失 $L^E(\theta, \phi)$

2)构造由上一步中的 3 个损失加权和构成的联合损失 $L(\theta, \phi)$;

3)根据联合损失,使用随机梯度下降法来更新模型参数;

4)循环步骤 1)–3)直到满足终止条件。

3 算例仿真分析

3.1 实验设置

选取 solomon 数据集中 R206、R207、C206、C207、RC206 和 RC207 等 6 组算例作为实验数据,每组算例中需求点均为 100。GMTVRPTW 模型的主要参数选取情况如表 2 所示。

3.2 不同类型算例的仿真结果分析

R206、R207 算例的运输时间窗大小为 10,运输规划窗口大小为 1000;C206、C207 算例的运输时间窗大小为 90,运输规划窗口大小为 3390。各算例的最优配送方案路线图如图 2 至图 7 所示,不同算例的仿真结果如表 3 所示:

表 2 实验参数设置

参数名称及符号	参数设置	参数名称及符号	参数设置
车辆行驶速度 v^*	71km/h	车辆使用时间成本 α	85 元/h
单位时间惩罚成本 PC	1min	雇佣司机在单位时间成本 φ	30 元/h
车辆在路径 d_{ij} 中的油耗率 f_{ij}	0.15kg/km	车辆的固定发车成本 θ	300 元
车辆在 d_{ij} 的碳排放率 c_{ij}	0.3703L/kg	单位油耗成本 ρ_f	8.0 元/L
单位碳排放成本 ρ_f	0.538 元/kg	控制系数 λ	0.015
惩罚系数 β	1	迭代次数 ω	500

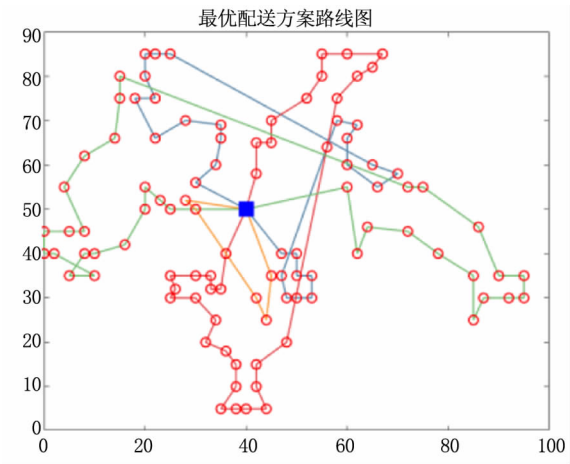


图 2 C206 算例最优配送方案图

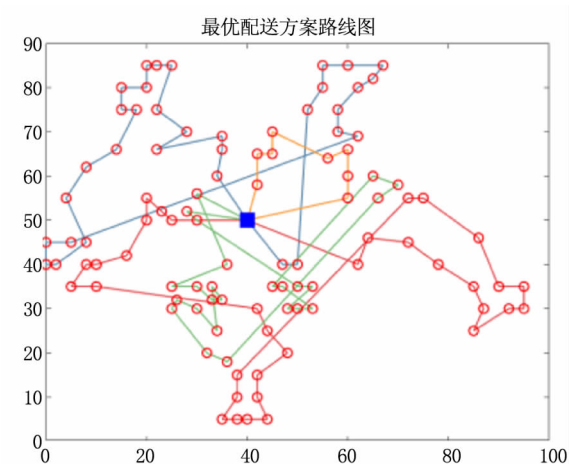


图 3 C207 算例最优配送方案图

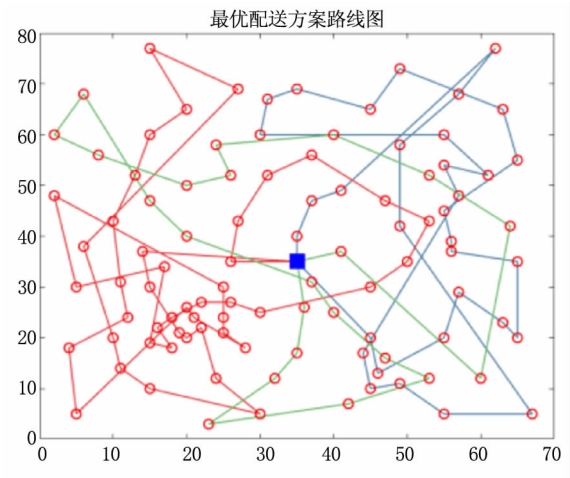


图 4 R206 算例最优配送方案图

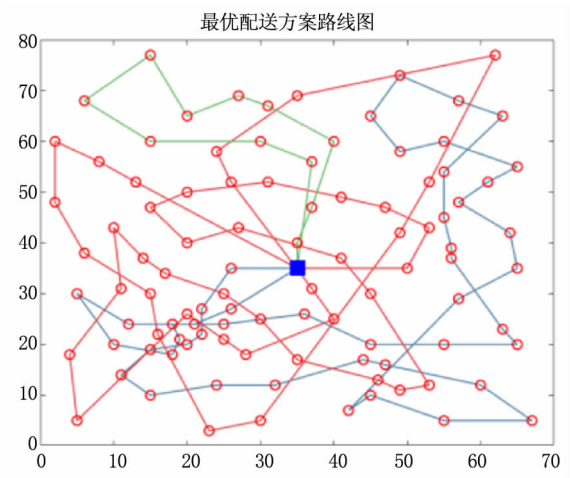


图 5 R207 算例最优配送方案图

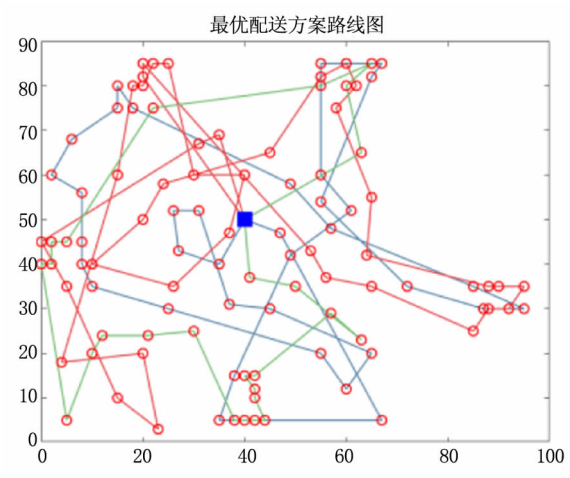


图 6 RC206 算例最优配送方案图

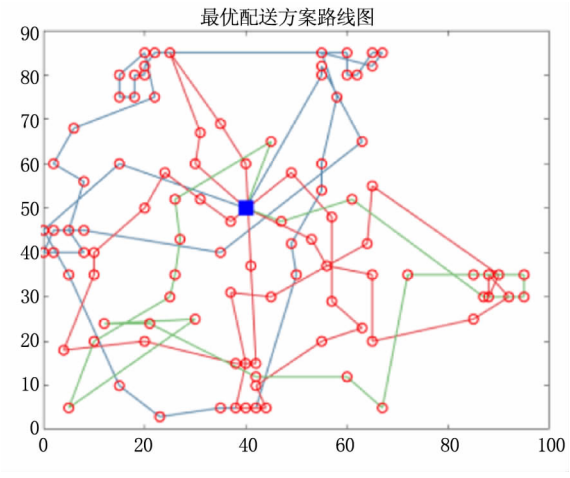


图 7 RC207 算例最优配送方案图

表 3 不同类型算例的仿真结果

数据集	车辆数/辆	总距离/km	总成本/元
R206	3	1182.79	8090.32
R207	3	1075.88	7440.35
C206	4	828.68	7579.87
C207	4	730.25	6822.06
RC206	3	1492.42	9972.55
RC207	3	1325.61	8958.49

由图 2 至图 5 的最优配送方案图与表 3 可知:
1)对于不同的算例,模型使用车辆数量不完全相同,这是因为各算例中每个客户所要求的时间窗和

服务时长不尽相同,模型需要兼顾运输成本和用户体验;2)R206 和 R207、RC206 和 RC207(R 组)所需要车辆数量少于 C206 和 C207(C 组)的车辆数量,其原因是 R 组客户点分布比 C 组更加分散,方便车辆返回物流中心实现路径串联。3)R 组算例效果比 C 组算例效果差,这是因为 C 组算例比 R 组算例有着更宽裕的时间窗口。

3.3 不同优化目标的仿真结果分析

为比较不同优化目标对模型效果的影响,分别将公式(5)所示模型的总成本目标与单独考虑油耗和碳排放成本、车辆行驶总距离的优化目标进行对比分析,在 R206 算例上实验,利用 Actor-Critic 深度强化学习算法求解,结果如表 4 所示。

表 4 不同优化目标的仿真结果比较

优化目标	总成本/元	油耗和碳排放成本/元	行驶距离/km	车辆使用成本/元	车辆数/辆
油耗和碳排放成本	12468.33	2125.65	1411.68	10342.68	6
行驶总距离	12035.65	2062.52	1376.39	9973.10	4
总成本	8090.32	2241.61	1182.79	5848.71	3

根据表 4 分析,1)相对于追求单一优化目标,综合考虑油耗、碳排放、车辆运营、驾驶员雇佣与客户满意度等多因素运输总成本最小的优化目标在碳排放与油耗费用、车辆行驶总距离和总成本方面展现更低的数值;2)相较仅关注单一行驶距离、油耗与碳排放成本的优化目标,GMTVRPTW 模型在考虑多种因素情况下实现了车辆使用效率和车辆数量配置的最优化,实现多种效益的最大化。

3.4 不同时间窗类型的仿真结果分析

分别采用软时间窗和硬时间窗约束的 GMTVRPTW 模型,使用基于 Actor-Critic 算法对模型求解,分析求解时间、车辆数以及车辆运输综合成本,分析对比表 5 如下所示:

根据表 5 的分析:1)相对于硬时间窗约束的 GMTVRPTW 模型,软时间窗约束的模型可以减少运输过程中所需的车辆数;2)根据运输总成本结果,带软时间窗的 GMTVRPTW 模型实现了最低成本运输的方案,软时间窗口约束更加贴合多行程车辆路径规划且能更好地兼顾客户服务满意度,实现运输成本和客户服务满意度效益的双向优化。

3.5 不同求解算法的仿真结果分析

基于 Actor-Critic 深度强化学习算法与工程中常用的蚁群方法对 GMTVRPTW 模型进行求解,其对比表 6 如下所示:

由表 6 可知:1)利用基于 Actor-Critic 方法与蚁群方法对 6 组数据集实验,均可求得优秀解,但 Actor-Critic 算法求解可以有效减少运输车辆数量,且明显缩短了求解时间;2)在车辆运输成本上,Actor-Critic 方法相较于蚁群算法实现的最低成本解决方案,显现出了更优的经济性。

表 5 软/硬时间窗下 GMTVRPTW 的最优运输方案综合成本比较

数据集	软时间窗		硬时间窗	
	总成本/元	车辆数/辆	总成本/元	车辆数/辆
R206	8090.32	3	11653.86	14
R207	7440.35	3	10732.62	14
C206	7579.87	4	9326.28	9
C207	6822.06	4	11826.75	9
RC206	9972.55	3	13536.32	14
RC207	8958.49	3	13271.01	14

表 6 不同算法模型求解

数据集	基于 Actor-Critic 深度强化学习算法			蚁群算法		
	求解时间/s	总成本/元	车辆数	求解时间/s	总成本/元	车辆数
R206	44.75	8090.32	3	156.69	11653.86	14
R207	41.72	7440.35	3	154.95	10732.62	14
C206	53.36	7579.87	4	209.11	9326.28	9
C207	54.56	6822.06	4	204.03	11826.75	9
RC206	45.56	9972.55	3	164.23	11571.35	8
RC207	44.36	8958.49	3	163.97	11324.11	8

4 结 论

货物物流是国内碳排放的主要来源之一,为了实现降低物流运输过程中碳排放和运输成本最低的目的,本文综合考虑碳排放成本、车辆使用时间成本、车辆固定成本、司机的雇佣成本和惩罚成本等因素,以及在配送过程中车辆距离、载重、速度等多种因素对油耗和碳排放的影响,针对客户对配送过程中提前和延迟完成配送服务带来的时间差异与服务满意度,通过引入惩罚成本刻画软时间窗,构建带软时间窗口的多行程绿色车辆路径优化模型 GMTVRPTW。本研究采用 solomon 数据集 中的算例利用 Actor-Critic 算法对 GMTVRPTW 模型进行验证。实验结果表明,构建的模型在最小碳排放成本情况下,可以对物流车辆的路径进行优化,有效地减少了碳排放,实现了油耗成本、车辆使用成本等在内的物流总成本的最小化目标,为企业绿色配送和碳减排提供决策参考,提高物流企业的市场核心竞争力。

参考文献

[1] BRUGLIERI M, MANCINI S, PEZZELLA F, et al. A path-based solution approach for the green vehicle routing problem[J]. Computers & Operations Research, 2019, 103: 109—122.

[2] 张静怡,司夏萌,张虹. 基于低碳排放的应急物流车辆路径优化[J]. 物流技术,2023,42(11):59—67.

[3] 张金良,李超. 碳排放影响下的动态配送车辆路径优化研究[J]. 中国管理科学,2022,30(9):184—194.

[4] 刘虹,傅晓敏. 考虑同时取送随机需求的多行程车辆路径研究[J]. 西安电子科技大学学报(社会科学版),2019,29(3): 87—95.

[5] 范厚明,吴嘉鑫,耿静,等. 模糊需求与时间窗的车辆路径问

题及混合遗传算法求解[J]. 系统管理学报,2020,29(1): 107—118.

[6] TAŞ D, DELLAERT N, VAN WOENSEL T, et al. Vehicle routing problem with stochastic travel times including soft time windows and service costs[J]. Computers & Operations Research, 2013, 40(1): 214—224.

[7] HE M, WEI Z, WU X, et al. An adaptive variable neighborhood search ant colony algorithm for vehicle routing problem with soft time windows[J]. IEEE Access, 2021, 9: 21258—21266.

[8] 周鲜成,余玮瑛,李松明,等. 考虑时间依赖和同时取送货的双目标绿色车辆路径问题研究[J]. 系统科学与数学,2024, 44(9):2798—2815.

[9] ESKANDARPOUR M, OUELHADJ D, HATAMI S, et al. Enhanced multi-directional local search for the bi-objective heterogeneous vehicle routing problem with multiple driving ranges[J]. European Journal of Operational Research, 2019, 277(2): 479—491.

[10] 周雅兰,廖易天,栗筱,等. 深度强化学习 Memetic 算法求解取送车辆路径问题[J]. 计算机科学与探索,2024,18(3): 818—830.

[11] 罗佳,李朝锋. 基于残差图卷积网络与深度强化学习的需求可拆分车辆路径优化算法[J]. 控制理论与应用,2024,41 (6):1123—1136.

[12] XIN L, SONG W, CAO Z, et al. Multi-decoder attention model with embedding glimpse for solving vehicle routing problems[C]. Proceedings of the AAAI Conference on Artificial Intelligence Press,2021, 12042—12049.

[13] RAAJAN J, SRIHARI P V, SATYA J P, et al. Real time path planning of robot using deep reinforcement learning [J]. IFAC—PapersOnLine, 2020, 53(2): 15602—15607.

[14] HICKMAN A J . MEET: Methodology for calculating transport emissions and energy consumption[M]. Office for Official Publications of the European Communities, 1999.

[15] 李进,张江华. 基于碳排放与速度优化的带时间窗车辆路径问题[J]. 系统工程理论与实践,2014,34(12):3063—3072.

[16] PALMER A . The development of an integrated routing and carbon dioxide emissions model for goods vehicles [D]. Cranfield University, 2007.