

基于DRL与元启发式算法的冷链物流路径优化研究

赵墨然, 周艳聪*, 周钰京, 冯大中, 廖欧珣, 洪宇恒

天津商业大学信息工程学院, 天津

收稿日期: 2025年9月20日; 录用日期: 2025年10月10日; 发布日期: 2025年10月22日

摘要

随着冷链物流市场的快速扩容与智能化升级, 多约束路径优化问题成为行业发展的关键瓶颈。深度强化学习(Deep Reinforce Learning, DRL)与元启发式算法相融合的方法将二者优势结合起来, 在路径优化问题中展现巨大潜力。据此, 本文分析了不同融合模式的设计逻辑差异, 探讨了各类方法的优势与局限性, 并展望了未来研究方向。

关键词

冷链物流, 路径优化, 深度强化学习, 元启发式算法, 算法融合

Research on Cold Chain Logistics Route Optimization Based on DRL and Meta-Heuristic Algorithm

Moran Zhao, Yancong Zhou*, Yujing Zhou, Dazhong Feng, Ouxun Liao, Yuheng Hong

School of Information Engineering, Tianjin University of Commerce, Tianjin

Received: September 20, 2025; accepted: October 10, 2025; published: October 22, 2025

Abstract

With the rapid expansion and intelligent upgrading of the cold chain logistics market, the problem of multi-constraint path optimization has become a key bottleneck in the development of the industry. The fusion of deep reinforcement learning (DRL) and meta-heuristic algorithms combines the

*通讯作者。

文章引用: 赵墨然, 周艳聪, 周钰京, 冯大中, 廖欧珣, 洪宇恒. 基于 DRL 与元启发式算法的冷链物流路径优化研究[J]. 管理科学与工程, 2025, 14(6): 1020-1024. DOI: 10.12677/mse.2025.146119

advantages of the two and shows great potential in path optimization problems. Based on this, this paper analyzes the differences in design logic among various fusion models, discusses the advantages and limitations of various methods, and looks forward to future research directions.

Keywords

Cold Chain Logistics, Path Optimization, Deep Reinforcement Learning, Meta-Heuristic Algorithm, Algorithms Fusion

Copyright © 2025 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

随着“双循环”战略的推动,我国冷链需求激增,冷链物流行业面临巨大挑战。传统优化方法在多约束下存在解质量冗杂和迭代效率低的问题,这不仅增加了配送时长,也降低了客户满意度。而深度强化学习(Deep Reinforce Learning, DRL)与元启发式算法相融合的方法因其兼具动态决策和全局寻优的优势,为突破这一困境提供了关键方向。

2. 相关研究综述

2.1. 冷链物流路径优化研究现状

我国作为一个能源消耗和碳排放的大国,发展绿色低碳的物流运输模式迫在眉睫,而车辆路径问题的研究正是实现节能减排和绿色运输的有效途径之一[1]。因此,不少学者对此进行了研究。方文婷等人基于绿色物流理念,构建了融入绿色成本的路径优化模型,将A*算法和蚁群算法相结合形成混合蚁群算法并通过仿真实例验证其有效性[2]。黄继磊在时间和载重约束下构建融合惩罚函数的目标函数,并建立多层多分支冷链网络架构,通过交叉与变异算子来改进粒子群算法,经实验表明,所提方法在路径优化上效果佳[3]。张天瑞等人针对交通拥堵导致冷链配送时效下降的问题构建了包含拥堵约束的生鲜冷链配送模型,并且运用改进的哈里斯鹰算法进行求解,最后,通过仿真实验验证其模型和算法的有效性[4]。江云倩等人构建了配送成本最小化模型,采用随机遍历抽样并引入自适应的交叉变异算子,另外采用精英保留策略保护最优解,形成改进的自适应遗传算法,最后通过与优化前的总配送成本和碳排放进行对比验证其显著效果[5]。针对冷链物流配送成本高、车辆利用率低等问题,李军涛等人构建了多目标的多车型路径优化模型,采用了自适应遗传模拟退火算法进行求解,结果显示,多车型配送综合优势显著且该算法性能优于传统遗传算法[6]。吴暖等人对于客户高时间服务需求,提出了时间容忍度的概念及其量化方法,结合节约法、精英策略等设计了改进多目标模拟植物生长算法,最后经冷链企业实例及与非支配排序遗传算法 II (Non-dominated Sorting Genetic Algorithm II, NSGA-II)对比验证其算法优越性[7]。

2.2. 基于 DRL 和元启发式算法的冷链物流路径优化研究现状

为实现冷链路径的高效优化,学者们尝试过遗传、蚁群等多种元启发式算法,也探索过通过深度强化学习(Deep Reinforce Learning, DRL)来解决问题。但由于冷链场景的多约束与实时动态性,单一方法渐显乏力,由此,催生出基于两者融合的解决方法。同时,结合 DRL 与元启发式算法在冷链物流路径优化中的融合逻辑及场景适配特性,笔者将其归纳为以下三种类型:

1) 深度强化学习主导 - 元启发式算法辅助

该类研究以深度强化学习控制全局路径, 元启发式算法对局部进行优化求解。Kool 等人提出了一个基于注意力机制的深度强化学习模型, 其通过深度强化学习直接输出路径问题的解, 再通过局部搜索进行优化, 得出最终结果[8]。Ma 等人先是输入车辆路径问题的基础数据, 强化学习输出全局路径, 然后采用 K-opt 的局部搜索策略通过交换顺序对全局路径进行微调, 输出最终结果[9]。王万良等人构建了多智能体深度强化学习模型, 该模型先定义了强化学习核心要素, 然后借助注意力机制搭建策略网络, 同时结合 2-opt 和采样搜索优化解, 最终可快速输出高质量路径[10]。Silva 等人针对路径规划和调度问题, 提出自适应多智能体框架, 其核心是通过强化学习驱动多智能体分布式探索, 再融合元启发式算法优化解从而得出结果[11]。综上所述, 以上研究均是深度强化学习进行全局动态决策, 元启发式算法进行局部优化。

2) 元启发式主导 - 深度强化学习辅助

该类研究中元启发式主导全局搜索, 深度强化学习辅助生成初始解或优化参数等, 从而得到最终结果。Ye 等人提出深度蚁群算法(Deep Ant Colony Optimization, DeepACO)框架, 其中图神经网络自动学习启发式规则, 之后, 训练好的图神经网络模型辅助蚁群算法(Ant Colony Optimization, ACO), 二者信息相结合从而得出最终解[12]。Rodríguez-Esparza 等人提出了一种结合自适应的模拟退火算法(Simulated Annealing, SA)和多臂赌博机(Multi-Armed Bandits, MAB)的超启发式算法, 其中自适应模拟退火算法为主搜索框架, 负责接受或拒绝解, 多臂赌博机按照规则历史收益动态择优进行求解[13]。综上, 上述研究均以深度强化学习辅助元启发式得出优化方案。

3) 深度强化学习 - 元启发式动态交互

该类型中深度学习算法和元启发式进行双向迭代, 形成闭环。许波等人的专利中, 通过 Q 学习机制实时动态选择多种邻域操作, 另外, 依据奖励反馈自适应调整蚁群搜索策略, 从而形成闭环优化[14]。Leng 等人构建了一个基于 Q 学习的超启发式框架, 该框架通过设计状态空间感知实时路径约束, 并且依据奖励信号来动态选择底层领域的启发式规则, 实现了在求解过程中根据环境反馈自适应调整搜索行为的闭环交互机制[15]。综上所述, 以上研究均通过双向动态交互得出结果。

虽然上述三类划分能从“控制流”维度梳理技术融合的逻辑, 但在具体研究场景中, 三类框架存在边界模糊和重叠的问题。例如: Ye 等人的研究中, 虽然蚁群算法主导全局路径的信息素更新, 深度强化学习辅助局部节点的选择, 但是二者之间也存在双向反馈, 这既满足元启发式算法主导 - 深度强化学习辅助, 又体现出二者动态交互, 这导致了类别边界模糊[12]; 另外, Wu 等人的研究中存在前期以权重感知深度强化学习为核心来生成初始解, 随后转为以 NSGA-II 为核心优化解的质量的现象[16]。该现象模糊了单一主导还是双向交互的类别边界[15]。

尽管该分类方法存在边界模糊的问题, 但选择控制流作为分类维度可以更精准地明确两者融合的协作内核, 并且可以直击两类算法协作的关键逻辑, 这相较于其他潜在维度, 控制流既能反映架构设计的核心特征, 又能为对于适配场景的需求进行算法协作模式的调整提供理论指引, 由此为深度强化学习与元启发式算法融合的深入研究筑牢了的理论根基。

3. 当前研究的局限性

1) 深度强化学习主导 - 元启发式算法辅助

深度强化学习主导 - 元启发式算法辅助的策略使元启发式信息与深度强化学习策略耦合过紧的场景下策略退化。Ma 等人在其研究中构建了以深度学习模型为主导并深度内嵌传统优化算子的框架, 该紧密耦合的设计虽然在特定训练场景下表现高效, 但由于策略刚性及过度依赖于内嵌算子, 导致了在问题特

征变化时出现的适应性不足与性能显著衰退的现象[17]。另外，超参数维数激增。Ye 等人的研究中，DeepACO 除了将神经网络的训练超参数与蚁群算法的传统控制参数相叠加，还引入了新的信息素初始化强度等协同参数，这导致整体超参数空间维度显著扩张[12]。

2) 元启发式主导 - 深度强化学习辅助

元启发式主导 - 深度强化学习辅助的策略存在实时性响应瓶颈。Wang 等人在其论文中明确提及：高维状态会增加 DRL 的计算延迟、离线训练难适配极端动态场景，另外，自适应大邻域搜索(Adaptive Large Neighborhood Search, ALNS)与 DRL 存在串行依赖，这会导致实时性响应不足的问题[18]。其次，奖励稀疏与延迟。张景玲等人的研究中，作为高层策略的深度 Q 网络，仅在底层元启发式算子完成整个路径的解的构造与评估后，才能获得一个稀疏的全局奖励信号，这使得学习效率严重被制约[19]。

3) 深度强化学习 - 元启发式动态交互

深度强化学习 - 元启发式动态交互策略使得探索与利用失衡。Li 等人的研究中将遗传算法(Genetic Algorithm, GA)与 Rainbow DRL (Rainbow: Combining Improvements in Deep Reinforcement Learning)结合，GA 全局探索网络结构与 DRL 局部利用经验冲突，这导致了模型早期快收敛但后期性能停滞、泛化不足的现象[20]。另外，两者融合后的策略可解释性不足。Leng 等人在其研究中采用 Q 学习动态选择底层启发式规则，但其决策逻辑被编码在不断更新的 Q 值函数中，从而造成难以追溯特定路径决策的依据的结果[15]。

综上所述，当前研究中仍存在着元启发式信息与深度强化学习策略耦合过紧的场景下策略退化、超参数维数激增、实时性响应瓶颈、奖励稀疏与延迟、探索与利用失衡、深度强化学习和元启发式算法融合后的策略可解释性不足等问题，仍需要不断改进策略和方法，以期更适应实际场景，得到更优方案。

4. 未来展望

当前，在深度强化学习与元启发式算法融合的研究领域中，现有成果多聚焦于理论架构构建及小规模仿真验证，缺乏大规模真实场景的验证；对于超参数维数激增、奖励稀疏与延迟等的系统探究也仍有欠缺。而实际冷链情景规模大、动态变化且系统对实时响应和决策透明性存在现实需求，因此，深度强化学习与元启发式算法的融合框架需要在实时响应能力与决策透明性上寻求突破，具体而言可重点围绕以下研究问题展开深入研究：

1) 当冷链配送过程中突发订单插入、配送车辆故障等动态实时问题时，深度强化学习 - 元启发式动态交互的模式与深度强化学习主导 - 元启发式辅助的模式相比，在解的质量和计算效率中存在怎样的权衡？

2) 如何构建适配于深度强化学习与元启发式融合架构的决策透明化模块，以清晰体现两类算法在冷链路径优化中的分工情况与关键信息传递的细节，从而帮助物流企业能够快速定位决策偏差的根源，提升融合框架在实际应用中的可信度？

3) 如何设计深度强化学习与元启发式算法融合的架构，在包含多温层需求的大规模冷链配送网络中，能够实现运输成本显著降低，碳排放有效削减与客户满意度良好的多目标协同优化，同时又能确保架构适配于各种复杂场景？

通过上述问题的探索，期望填补深度强化学习与元启发式算法融合在冷链场景的研究空白，为企业提供适配的技术方案，助力低碳冷链的高质量发展。

基金项目

本文由国家级大学生创新创业项目(202510069028)资助。2025 年全国高校、职业院校物流教改教研

立项课题(课题号: JZW2025036)。

参考文献

- [1] 李卫朝. 基于深度强化学习的动态多车型绿色车辆路径优化研究[D]: [硕士学位论文]. 西安: 西安电子科技大学, 2022.
- [2] 方文婷, 艾时钟, 王晴, 等. 基于混合蚁群算法的冷链物流配送路径优化研究[J]. 中国管理科学, 2019, 27(11): 107-115.
- [3] 黄继磊. 基于粒子群自进化的冷链物流运输路径优化方法[J]. 包装工程, 2025, 46(11): 285-293.
- [4] 张天瑞, 祝芳芳, 牛慧媛. 基于改进哈里斯鹰算法的生鲜品冷链物流路径优化研究[J]. 重庆师范大学学报(自然科学版), 2025, 42(2): 1-13.
- [5] 江云倩, 杨慧敏, 彭程, 等. 考虑碳排放和时间窗的冷链物流配送路径优化研究[J]. 包装工程, 2024, 45(3): 262-268.
- [6] 李军涛, 刘明月, 刘朋飞. 生鲜农产品多车型冷链物流车辆路径优化[J]. 中国农业大学学报, 2021, 26(7): 115-123.
- [7] 吴暖, 代焕杰, 李季涛, 等. 考虑时间容忍度的冷链物流配送路径多目标优化[J]. 交通运输系统工程与信息, 2023, 23(2): 275-284.
- [8] Kool, W., van Hoof, H. and Welling, M. (2018) Attention, Learn to Solve Routing Problems! arXiv: 1803.08475.
- [9] Ma, Y., Liu, Q. and Tang, J. (2023) Learning to Search Feasible and Infeasible Regions of Routing Problems with Flexible Neural k-Opt. arXiv: 2310.18264.
- [10] 王万良, 陈浩立, 李国庆, 等. 基于深度强化学习的多配送中心车辆路径规划[J]. 控制与决策, 2022, 37(8): 2101-2109.
- [11] Silva, M.A.L., Souza, S.R.de, Souza, M.J.F. and Bazzan, A.L.C. (2019) A Reinforcement Learning-Based Multi-Agent Framework Applied for Solving Routing and Scheduling Problems. *Expert Systems with Applications*, **131**, 148-171.
- [12] Ye, H., Wang, J., Cao, Z., Liang, H. and Li, Y. (2023) DeepACO: Neural-Enhanced Ant Systems for Combinatorial Optimization. arXiv: 2309.14032.
- [13] Rodríguez-Esparza, E., Masegosa, A.D., Oliva, D. and Onieva, E. (2024) A New Hyper-Heuristic Based on Adaptive Simulated Annealing and Reinforcement Learning for the Capacitated Electric Vehicle Routing Problem. *Expert Systems with Applications*, **252**, Article ID: 124197. <https://doi.org/10.1016/j.eswa.2024.124197>
- [14] 许波, 姜艺, 刘玉敏, 等. 用于冷链食品运输过程的结合 Q 学习机制的变邻域蚁群优化调度方法[P]. 中国专利, CN115063177B. 2024-03-08.
- [15] Leng, L., Jin, Q., Chen, T., Wan, A. and Wang, Z. (2024) Energy-Conserving Cold Chain with Ambient Temperature, Path Flexibility, and Hybrid Fleet: Formulation and Heuristic Approach. *International Journal of Production Research*, **63**, 26-50. <https://doi.org/10.1080/00207543.2024.2355325>
- [16] Wu, R., Wang, R., Hao, J., Wu, Q., Wang, P. and Niyato, D. (2025) Multiobjective Vehicle Routing Optimization with Time Windows: A Hybrid Approach Using Deep Reinforcement Learning and Nsga-ii. *IEEE Transactions on Intelligent Transportation Systems*, **26**, 4032-4047. <https://doi.org/10.1109/tits.2024.3515997>
- [17] Ma, Y., Li, J., Cao, Z., Song, W., Zhang, L., Chen, Z. and Tang, J. (2021) Learning to Iteratively Solve Routing Problems with Dual-Aspect Collaborative Transformer. *Advances in Neural Information Processing Systems*, **34**, 11096-11107.
- [18] Zhao, J., Mao, M., Zhao, X. and Zou, J. (2021) A Hybrid of Deep Reinforcement Learning and Local Search for the Vehicle Routing Problems. *IEEE Transactions on Intelligent Transportation Systems*, **22**, 7208-7218.
- [19] 张景玲, 冯勤炳, 赵燕伟, 等. 基于强化学习的超启发式求解有容量车辆路径问题[J]. 计算机集成制造系统, 2020, 26(4): 1118-1129.
- [20] Li, P., Zheng, Y., Tang, H., Fu, X. and Hao, J. (2024) EvoRainbow: Combining Improvements in Evolutionary Reinforcement Learning for Policy Search. *Proceedings of the 41st International Conference on Machine Learning*, **235**, 29427-29447.