Dean&Francis

ISSN 2959-6157

Research on the Application of Deep Reinforcement Learning in Logistics Scheduling

Changsheng Lu

Department of Logistics Engineering, Nanjing Forestry University, Jiangning District, Nanjing, Jiangsu Province, China, Postal Code: 211100
Corresponding author: Changsheng Lu
Email: 3193930688@qq.com

Abstract:

Traditionally, the approaches of logistics scheduling are hard to adapt to the rapid changes of the operational conditions and many of them do not have self-adaptive optimization capability, which will leads to the unsmooths and high operational expenses. In this research, we use Deep Reinforcement Learning (DRL) to improve the logistics scheduling. We take the warehouse order picking as an application example, we build the simulated warehouse environment on a grid system and design the intelligent scheduling based on Proximal Policy Optimization (PPO) algorithm. We compare the performance of it with the usual A* path planning method. The results show that the DRL model built on PPO has better results than the A* algorithm for several important measures, for example, the average steps to finish, the rate of order finish and the total reward accumulation. This investigation proves the possibility of DRL to make flexible and smart logistics scheduling, which is both theoretical reference and .a low-cost simulation tool for the intelligent transformation of logistics enterprises.

Keywords:. Deep Reinforcement Learning; Logistics Scheduling; Order Picking; PPO Algorithm; A* Algorithm

1. Introduction

With the expansion of e-commerce and new retail models, the development of the logistics sector has been accelerated. With that, the requirements for the scheduling operations are to be faster, more flexible and more intelligent. The approaches of the scheduling that we have until now are mostly based on rules that are fixed and on models based on experience. They are not adequate to assume changes in the volume of the orders, in the complex configuration of the warehouses or on the assignment of the deliveries with several constraints. These classic techniques are not able to fast adapt to change of conditions, to reconcile different goals in the operations, for example the speed of the operations and the cost, or to improve by itself its approach with the knowing of the operations done in the past. So, the in-efficience on the logistics process are still a problem.

It is a kind of merging of the perception ability of deep learning and the benefit of decision making of reinforcement learning, deep reinforcement learning. The agent trained with DRL by doing the continuous interaction with the environment can acquire the good scheduling strategy adaptively. The method has proven to solve the complex sequential decision making problems, such as the path planning of robot and the optimization of the game strategy, so that it can be a possible way to break through the limitation of the traditional scheduling method in the dynamic logistics situation.

This paper focuses on the warehouse order picking, which is one of the logistic operations. The main goals are to design a smart scheduling scheme by deep reinforcement learning(DRL) to improve the operation performance, to deal with the multi-objective optimization with given constraints, and to show the advantage of the proposed method compared with the usual method. The main contributions of this work are in two parts: First, a clear and reproducible grid-based simulation framework for the warehouses is built, which transforms the real scheduling problem into some standard reinforcement learning problems; then, the introduced model can provide the logistics company a simulation environment in cheap and safe way to test the scheduling strategy based on DRL, which can help the logistics company to turn towards the intelligent operation.

2. Literature Review

The use of deep reinforcement learning for logistics scheduling has received much interest of both researchers and practitioners in recent times. The works done so far have considered a variety of contexts like last mile delivery routing, management of the automated guided vehicle (AGV) in the warehouse, or the synchronization of the supply chain. Many results of these works have already been used in the operation environment.

Work done in all over the world provides us with a solid theoretical basis for the use of DRL to complex decision making problems. For example the contributions of Mykel J. Kochenderfer regarding dynamic decision making problems, say in air traffic management, give us methodological perspectives that can be transferred to the optimization of logistics routes or of immediate schedule tasks. Work done by Yoshua Bengio about how to tackle combinatorial optimization problems by DRL has surpassed the performance limits inherent to the usual heuristic methods, and presents new ways for dealing with large logistics network schedules.

Research in the country also treats theory and industrial application equally well. The group headed by Yang Yu has applied multi-agent reinforcement learning to large logistics scheduling framework. The platform which is made available by them in public, Tianshou, has become one of the basic stuffs for works in this area. Feng Gengzhong and Liu Yuewen have proposed improved deep reinforcement learning techniques for the routing problems for different types of vehicles, which have the good capabilities for handling complex constraints.

However, even with all these developments, there still is the need for some further investigation in some directions, e.g. the improvement of the adaptive algorithms, the application of these methods in practice in complicated real-world situation, and the trade-offs in the multi-objective optimization. In this paper we further the existing work by tackling those open questions with some specific experimental investigations in the context of warehouse order picking operations.

3. Methodology

3.1 Experimental Environment Setup

All the tests are done on a system with the operating system Windows 10. We use the Python version 3.9. The necessary software packages are: gymnasium for simulation of the warehouse environment, Stable-Baselines3 for applying the method PPO, NumPy for the numerical data, and Matplotlib for the charts of the results.

The warehouse that is simulated is based on a 10 by 10 grid, with storage racks, products, a robotic picker, a fixed dispatch location, and generated order assignments. There are 15 rack locations placed at random on the grid, and each of these locations has a different product category. Each order assignment has 3 to 5 different products, and the robotic agent has to collect all the products that are specified in the order and bring them to the fixed dispatch location.

3.2 Algorithm Design

3.2.1 A* Algorithm

The baseline method is the A* search algorithm with a nearest-neighbor heuristic. The robotic agent chooses the nearest rack which has not been visited yet as next goal and finds the best way via A*. It goes like this until every product is picked up and sent, also the system has no self learning.

3.2.2 PPO Algorithm

The Deep Reinforcement Learning framework is built with the method of Proximal Policy Optimization. The

ISSN 2959-6157

state is the representation of the position of the robotic agent and the position of the delivery location. The action space is a set of five different, discrete actions: up action, down action, left action, right action and the object pickup action.

We set up a composite reward mechanism to steer the learning. The agent is rewarded with +10 for performing a pick action, +0.1 for reducing the Euclidean distance to the target location, +50 for finishing an entire order, and -1 for performing any action that is not valid or is leading to a collision. The episode finishes either when the order is fulfilled or when a maximum number of steps is reached.

3.3 Training and Testing Protocol

We train the PPO model for 1000 epochs, and each epoch contains 100 different order fulfillment tasks. After the model has converged, we run 500 independent order picking trials for both the A* algorithm and the trained PPO model in the same environment. The performance is measured by the mean number of steps of the done, the percentage of done orders, and the average cumulative reward.

4. Results and Analysis

4.1 Comparative Performance Evaluation

Table 1 displays the quantitative evaluation outcomes for the two methods under consideration.

Table 1 Performance Comparison of A* and PPO Algorithms

Evaluation Indicator	A* Algorithm	PPO Algorithm	Improvement
Average Completion Steps	89.2	56.7	36.4%
Order Completion Rate	92.4%	98.8%	6.4%
Average Reward Score	68.5	95.2	38.9%

The data shows that the PPO method is better than A* method on every measured criterion. It decreases average number of steps for doing the task by 36.4%, increases successful order doing rate by 6.4%, and increases mean reward value by 38.9%.

4.2 Analysis of Results

From the visual check of the step progress we can see that the A* method keeps a always high and constant step number. On the other hand, the PPO method slowly converges to the lower step level when the training epochs go forward. The bar graph of the completion rates shows that PPO escapes the local optimum traps and the dead end situations often met by the A* algorithm, which is a nearest-neighbor strategy. The reward progression graph proves that the PPO learns to own the ability to self-improve its decision policies, which is not in the ordinary algorithmic way.

The PPO model shows better abilities of the global route design and environmental adaptation, which can manage the non-deterministic barriers and the complicated setting more efficiently. But it has also some disadvantages, such as the early-stage convergence is slow and the computation is high in large setting.

5. Conclusion

The present research proposes the application of deep reinforcement learning in the context of logistics coordination, in particular for the retrieval of orders in the warehouse. The main results are the following:

1. Deep reinforcement learning is able to overcome the limitations of the usual ways of scheduling and will increase the productivity and the possibility of flexibilization of the changing context of the logistics.
2. It can be seen that the PPO method is much better than the A* algorithm in some of the important performance index, which demonstrates the advantage of using the PPO method to improve the intelligent scheduling.
3. The grid based simulation structure that we have created in this work provides a repeatable test environment for scholarly enquiring and an affordable instrument for the logistics companies for testing the DRL implementations.

The future efforts will be made to improve the PPO algorithm to make the convergence faster and the algorithm can be operated in real time. The scope of the reward function will be enlarged to have more objective metrics of cost and energy saving. The model will be tested and verified in the real warehouse.

Although this study verifies the superiority of the PPO-based deep reinforcement learning scheduling model in warehouse order picking, several limitations remain to be addressed. First, the simulated warehouse environment only adopts a fixed 10×10 grid scene with a single robotic picker, without considering multi-robot collaborative picking, dynamic order arrival flows and variable warehouse layout structures in real logistics scenarios, which reduces the generalizability of the experimental conclusions. Second, the reward function only covers picking distance, order completion and collision penalty, lacking quantitative indicators related to equipment energy consumption, labor cost and order waiting time, failing to fully realize comprehensive multi-objective optimization of logistics operation benefits. Third, the research only compares the PPO algorithm with the classic A* path planning method, and does not conduct cross-comparison with other mainstream reinforcement learning algorithms such as DQN and DDPG, so the overall advantage of the proposed model cannot be fully demonstrated. In addition, this paper only cites a limited number of existing literature, lacking sufficient support from the latest domestic and international research on intelligent warehouse scheduling. Subsequent research can optimize the above deficiencies by building multi-agent simulation environments, expanding multi-dimensional reward functions, adding more comparative algorithms and enriching relevant research literature.

References

[1] Schulman, J., Wolski, F., Dhariwal, P., et al. (2017) Proximal Policy Optimization Algorithms. arXiv preprint arXiv:1707.06347.

[2] Yu, Y., Zhou, Z.H. (2020) Research on the Theory and Methods of Multi-agent Reinforcement Learning. Science Press, Beijing.

[3] Feng, G.Z., Liu, Y.W. (2022) Dynamic embedding deep reinforcement learning for heterogeneous vehicle routing problems with unloading time constraints. Journal of Management Sciences in China, 25(5): 1-16.

[4] Kochenderfer, M.J., Wheeler, T.A., Wray, K.H. (2019) Algorithms for Optimization. MIT Press, Cambridge.

[5] Bengio, Y., Lecun, Y., Hinton, G. (2021) Deep Learning for AI. Communications of the ACM, 64(7): 58-65.

[6] Zhang H, Li M. AGV path optimization based on improved DQN algorithm in intelligent warehouse [J]. Computer Integrated Manufacturing Systems, 2021, 27 (8):2312-2321.

[7] Liu S, Wang T. Multi-objective order picking scheduling with deep reinforcement learning under dynamic order demand [J]. Journal of Systems Engineering, 2023, 38 (2):245-254.

[8] Han X, Chen L. A review of deep reinforcement learning applications in logistics operation optimization [J]. Control and Decision, 2022, 37 (11):2421-2432.

[9] Huang J, Zhao Y. Real-time scheduling optimization of warehouse picking robots based on DDPG [J]. Computer Engineering and Applications, 2023, 59 (14):267-275.

[10] Li W, Sun Q. Dynamic warehouse layout optimization combining PPO and digital twin technology [J]. Journal of Industrial Engineering and Engineering Management, 2024, 38 (1):102-110.

[11] Gao R, Wu J. Heuristic algorithm versus reinforcement learning for order picking path planning: A comparative study [J]. International Journal of Logistics Research and Applications, 2022, 25 (9):987-1004.

[12] Wang Z. Research progress of intelligent logistics scheduling driven by deep reinforcement learning [J]. Systems Engineering Theory & Practice, 2023, 43 (5):1289-1301.