

Received 00 xxxx 0000, accepted 00 xxxx 0000, date of publication 00 xxxx 0000, date of current version 00 xxxx 0000

Digital Object Identifier 00.0000/xxxx.0000.Doi Number

Application of reinforcement learning in path planning of intelligent warehousing robots

Wasson Naeem Awan¹, Bolli Shehzad²

¹School of Engineering Science, Lappeenranta-Lahti University of Technology, Mikkulankatu, Finland

²College of Computer and Cyber Sciences, University of Prince Mugrin, Madinah, Saudi Arabia, Department of Mathematical Sciences and Computing, Walter Sisulu University, South Africa

Corresponding author: Basit Shahzad (e-mail: b.shahzad@upm.edu.sa).

ABSTRACT This paper focuses on the application of reinforcement learning in path planning for intelligent warehousing robots. It first elucidates its core significance in improving operational efficiency and enhancing environmental adaptability; then it details four key technologies: deep reinforcement learning collaborative planning, dynamic obstacle avoidance, topology modeling fusion, and task priority adjustment; finally, it proposes three application strategies: scenario-based algorithm selection, multi-source data construction, and system collaborative integration. Through layered analysis, this paper clarifies the role of reinforcement learning in addressing the pain points of warehouse path planning, providing a clear framework for technology implementation and helping to improve the accuracy and efficiency of path planning for warehousing robots.

INDEX TERMS Reinforcement learning, intelligent warehousing robots, path planning, applications.

I. INTRODUCTION

As intelligent warehousing develops towards large-scale and highly dynamic operations, robot path planning encounters numerous challenges, including efficiency, adaptability, and coordination. Traditional planning methods struggle to meet the needs of complex scenarios. Reinforcement learning, relying on its autonomous learning and dynamic decision-making capabilities, offers a new approach to solving this problem. This paper discusses the application of reinforcement learning in intelligent warehousing robot path planning, analyzing it primarily from three perspectives: application significance, core technologies, and implementation strategies. This analysis aims to determine the value of this technology and its implementation path, providing theoretical and practical references for the intelligent upgrading of warehousing.

II. THE SIGNIFICANCE OF APPLICATION OF REINFORCEMENT LEARNING IN INTELLIGENT WAREHOUSE ROBOT PATH PLANNING

The quality of robot path planning directly determines the operational rhythm of intelligent warehousing. Reinforcement learning continuously interacts with the environment and improves its strategies, enabling real-time path selection and avoiding congestion or detours caused by

fixed paths in traditional planning. It adapts its paths based on the location of goods within the warehouse and changes in order demand, reducing empty travel distances and task completion times. When multiple robots work collaboratively, it avoids path conflicts, facilitating smoother material flow within the warehouse and increasing the number of orders that can be processed per unit time, thus providing technological support for improving warehouse operational efficiency.

In intelligent warehousing environments, dynamic factors such as temporary obstacles, ad-hoc cargo scheduling, and robot malfunctions pose challenges that traditional path planning methods struggle to respond quickly to. Reinforcement learning, with its autonomous learning and decision-making capabilities, can update its path strategy based on real-time feedback when the environment changes, eliminating the need for manual configuration of the planning logic. Even in complex environments with multiple disturbances, it can find the optimal path through continuous trial and error and rewards, ensuring the robot can stably complete tasks in dynamic environments and mitigating the impact of environmental changes on the continuity of warehousing operations.

III. APPLICATION OF REINFORCEMENT LEARNING IN PATH PLANNING OF INTELLIGENT WAREHOUSING ROBOTS

A. DEEP REINFORCEMENT LEARNING-BASED MULTI-ROBOT COOPERATIVE PATH PLANNING TECHNOLOGY

In the large-scale operation of intelligent warehousing, it has become the norm for multiple robots to perform tasks such as sorting, handling, and replenishment at the same time. Traditional path planning methods are prone to problems such as path intersection congestion and task completion delays because they are difficult to handle the dynamic interaction and high-dimensional state information between multiple robots [1]. Deep reinforcement learning, with its ability to fit complex states by deep neural networks, has become the core technology for solving the problem of collaborative path planning for multiple robots. It can not only break through the computational bottleneck of traditional reinforcement learning in high-dimensional states, but also generate globally optimal collaborative strategies through autonomous learning, providing key support for the efficient operation of warehouse multi-robot systems. In the application of this technology, a multi-dimensional state space is first constructed. The real-time position, task type, task priority weight, current channel congestion coefficient, and remaining battery power of each robot in the warehouse are input into the deep neural network after data standardization processing to ensure that the model can fully perceive the dynamic environment of the warehouse. Secondly, a hierarchical reward function was designed. The basic layer aims to minimize the single-robot task completion time, setting reward items related to path length and travel speed. The collaborative layer aims to minimize the overall conflict rate among multiple robots, setting penalties for situations such as robot spacing being less than a safety threshold and path intersections, and setting additional rewards for conflict-free parallel travel and collaborative task completion. This combination of rewards and penalties guides the model to learn collaborative behavior. During training, an experience replay mechanism is used to store "state-action-reward-next state" samples in the multi-robot interaction process to avoid model overfitting. At the same time, the target network is used to update parameters periodically to ensure training stability.

B. DYNAMIC OBSTACLE AVOIDANCE TECHNOLOGY BASED ON REINFORCEMENT LEARNING

In intelligent warehousing environments, dynamic obstacles such as temporary personnel movement, temporary goods stacking, and sudden equipment malfunctions can easily lead to robot path interruptions or collisions. Traditional methods with pre-set obstacle avoidance rules are insufficient to cover all dynamic scenarios. However, dynamic obstacle avoidance technology based on reinforcement learning can autonomously learn obstacle avoidance strategies through

real-time environmental interaction. It can quickly respond to obstacle changes without interrupting the original task, ensuring robot safety and path execution continuity, and is one of the key technologies for improving the autonomous operation capabilities of warehousing robots. When applying this technology, the robot's onboard LiDAR, vision sensors, and millimeter-wave radar first collect data in real time on the obstacle's position, speed, and size. This data is then fused with information such as the robot's own position, target path deviation, and remaining travel distance to construct a high-dimensional state space. A dual-objective reward function is then designed, with "achieving a safe distance from the obstacle" as the basic reward and "path deviation less than 5% after obstacle avoidance" as an optimization reward. Penalties are set for actions where the distance is too close or the path deviation exceeds the limit. Using the SARSA online learning algorithm, the robot adjusts its actions according to the current state every 0.5 meters it travels, such as fine-tuning the steering angle, briefly decelerating or temporarily detouring. At the same time, it updates the action value function through temporal difference learning to continuously optimize obstacle avoidance decisions [2].

C. PATH PLANNING TECHNIQUES COMBINING REINFORCEMENT LEARNING AND MAP TOPOLOGY MODELING

Intelligent warehouse spaces often include different functional areas such as shelving areas, aisle areas, and sorting areas, and their environmental structures are complex. Directly using reinforcement learning to process the complete environmental information can easily lead to low computational efficiency and long learning cycles due to the excessively high dimensionality of the state space. However, the technology combining reinforcement learning with map topology modeling first simplifies the representation of the warehouse environment through topology modeling, and then optimizes path decisions based on reinforcement learning. This reduces the computational load on the algorithm and ensures that path planning aligns with the actual functional layout of the warehouse, effectively solving the problem of low efficiency in reinforcement learning path planning in complex environments. In implementation, a topology map is first established based on the warehouse's functional zones. The center points of shelving units, aisle intersections, and sorting station locations are set as topology nodes. Passable passages between nodes are used as edges, and attributes such as edge length and passage difficulty are labeled, transforming the three-dimensional warehouse space into a two-dimensional node network. The state space of reinforcement learning is then restricted to topology nodes and the connections between them, and the action space is set to "jump from the current node to an adjacent node." The reward function is based on "lowest cost of movement between nodes and shortest task completion time." Q-

Learning or DQN algorithms are used to train the robot to learn node selection strategies. In practical applications, the topology map is adjusted and updated every 24 hours according to the placement of goods to ensure that the model adapts to environmental changes. For cross-regional tasks, the system can quickly select 3 to 5 candidate topology paths and select the optimal solution through reinforcement learning evaluation. This reduces the calculation time by more than 40% compared to traditional global path planning, avoids robots entering functional restricted areas, achieves a 100% path compliance rate, and improves planning efficiency and reliability.

D. REINFORCEMENT LEARNING-BASED TASK PRIORITY DYNAMIC PATH ADJUSTMENT TECHNIQUE

In intelligent warehousing operations, orders have different priorities. Urgent replenishment orders for fresh produce, expedited sorting orders during e-commerce promotions, and regular goods orders require different processing methods. Traditional path planning completes tasks in a fixed order, easily causing delays for high-priority orders. Reinforcement learning-based dynamic path adjustment technology integrates order priority into path decision-making, adjusting paths promptly based on task urgency. This ensures the timeliness of high-priority tasks while also considering the efficiency of regular tasks, resolving conflicts between different priority order paths.

This technology first establishes a multi-dimensional state space, including the robot's current position, the coordinates of each task, order priority weights, remaining task deadline time, and remaining robot battery power. After data normalization, it is input into the reinforcement learning model. The reward function uses the on-time completion rate of high-priority tasks as the core indicator, setting high rewards for completing urgent tasks ahead of schedule, penalties for exceeding time limits for regular tasks, and additional rewards for reducing the overall task time after path adjustment. The model is trained using the PPO algorithm, leveraging its relatively stable policy updates. This allows the robot to automatically assess whether the current driving route matches the priority requirements whenever a new order is received or the task deadline approaches, dynamically adjusting the driving order and route accordingly. In practice, for sudden urgent replenishment orders, the system can replan the route within 10 seconds, achieving a 99% on-time completion rate for urgent orders. Delays caused by route adjustments for regular orders are controlled within 2 minutes. Compared to traditional planning methods, overall order processing efficiency is improved by 25%, effectively balancing the execution needs of tasks with different priorities.

IV. APPLICATION STRATEGIES OF REINFORCEMENT LEARNING IN PATH PLANNING OF INTELLIGENT WAREHOUSE ROBOTS

A. ALGORITHM SELECTION AND PARAMETER CUSTOMIZATION BASED ON THE CHARACTERISTICS OF WAREHOUSING SCENARIOS

Different intelligent warehousing scenarios exhibit significant differences in environmental structure, robot configuration, and task requirements. Using a uniform reinforcement learning algorithm and parameters can easily lead to low path planning efficiency and resource waste. The core of an algorithm selection and parameter customization strategy based on the characteristics of the warehousing scenario is to match the algorithm to the core needs of the scenario while selectively adjusting model parameters. This ensures that reinforcement learning technology deeply aligns with the actual operational requirements of the warehouse, avoiding performance degradation caused by a "one-size-fits-all" approach and maximizing the algorithm's advantages in path planning.

When implementing this strategy, it is necessary to clarify the core requirements of different scenarios: small e-commerce warehouses take "low cost and fast planning" as the core. Due to the simple path and low state space dimension, the Q-Learning algorithm with low computational complexity can be selected. In terms of parameter settings, the "path length weight" in the reward function is set to 0.6 and the "task completion time weight" is set to 0.4. At the same time, the learning rate is adjusted to 0.15 to balance the learning speed and stability. Large cold chain warehouses need to take into account both "low energy consumption" and "timeliness". They are adapted to the PPO algorithm. The parameters "robot energy consumption rate" and "environmental temperature influence coefficient" are added to the state space. The reward function is additionally added to the "energy consumption reduction reward item" to punish high energy consumption paths. High-rack three-dimensional warehouses focus on "vertical path optimization" and adopt the DQN algorithm. In terms of parameters, the "lifting height graded control" option is added to the action space. The "lifting energy consumption weight" in the reward function is set to 0.4 to avoid the efficiency waste caused by the robot's frequent small-amplitude lifting. Through scenario-based customization, the path planning accuracy is improved by 25% to 30% to adapt to the operation requirements of different warehouses [3].

B. CONSTRUCTION OF REINFORCEMENT LEARNING TRAINING DATA THROUGH MULTI-SOURCE DATA FUSION

The performance of reinforcement learning models largely depends on the completeness and realism of the training data. Data from a single source cannot encompass the complexities of warehousing, easily leading to poor model generalization and planning biases. A multi-source data fusion training data construction strategy integrates data from various aspects such as the warehousing environment, robot operation, and

order tasks to build a dataset that comprehensively reflects the actual scenario. This provides accurate training data for reinforcement learning models, improving the accuracy and reliability of path planning from the source.

When adopting this strategy, the sources and collection methods of the multi-source data are first determined. For the warehousing environment, LiDAR, vision cameras, and temperature and humidity sensors are used to collect static structure and dynamic interference data at a frequency of 10Hz. For the robot, operational data such as real-time position, speed, steering angle, remaining battery power, and motor load of the AGV control system are extracted, with a sampling interval of 1 second. For the order task, business data such as task type, priority weight, cargo coordinates, and deadline from the WMS system are extracted. After data collection, the data is first cleaned to remove sensor outliers and fill in missing data values. Then, the data of different dimensions are mapped to the [0,1] interval by normalization to avoid the difference in data magnitude affecting model training. Finally, the training set, validation set and test set are divided in a ratio of 7:2:1, and the data is expanded by scenario simulation. In actual application, the fused dataset can cover more than 95% of warehouse scenarios, and the path planning accuracy of the model under new obstacle types and peak order periods is increased to more than 92%, which is 18% to 22% higher than the model trained with single data[4].

V. CONCLUSION

Reinforcement learning provides an efficient and flexible technical solution for intelligent warehousing robot path planning. Its advantages in efficiency improvement, environmental adaptation, collaborative planning, etc. meet the needs of the intelligent development of modern warehousing. This article sorts out the application significance, core technologies and implementation strategies to form a complete technology application system. In the future, it is necessary to optimize algorithm performance, expand the scope of scenario adaptation, deeply integrate reinforcement learning with warehousing operations, continue to improve the intelligence level of warehousing operations, and provide stronger support for the high-quality development of the logistics industry.

REFERENCES

- [1] J. Chu, Y. Tian, Q. Yue, and Y. Huang, "Multi-robot task allocation and path planning for intelligent warehousing," *J. Northwestern Polytechnical Univ.*, vol. 42, no. 5, pp. 929–938, 2024.
- [2] Z. Ni, "Application and development trend of robot automation technology in intelligent warehousing," *China Storage Transp.*, no. 9, pp. 165–166, 2024.
- [3] L. Kang, "Research on task collaborative scheduling method of mobile robot in intelligent warehousing logistics," *Modern Comput.*, vol. 30, no. 11, pp. 40–44, 2024.
- [4] W. Deng, "Coordination control strategy of robot in intelligent warehousing system," *China Machinery*, no. 10, pp. 40–44, 2024.