Molecular & Cellular Biomechanics 2025, 22(5), 1507.
https://doi.org/10.62617/mcb1507

Article

Deep reinforcement learning and biomechanical modeling are integrated to optimize the scheduling problem of intelligent logistics and warehousing robots

Shuzhao Dong

Institute of Logistics Science and Engineering, Shanghai Maritime University, Shanghai 201306, China; dongzhaoshu82@163.com

CITATION

Dong S. Deep reinforcement learning and biomechanical modeling are integrated to optimize the scheduling problem of intelligent logistics and warehousing robots. Molecular & Cellular Biomechanics. 2025; 22(5): 1507.
https://doi.org/10.62617/mcb1507

ARTICLE INFO

Received: 8 February 2025
Accepted: 6 March 2025
Available online: 24 March 2025

Abstract: This study introduces an innovative approach to optimizing the scheduling of intelligent logistics and warehousing robots by integrating deep reinforcement learning (DRL) with biomechanical modeling. Leveraging a comprehensive dataset from a large-scale logistics company, the research formulates the scheduling problem as a Markov Decision Process (MDP) and incorporates biomechanical principles to accurately model robot energy consumption. A Deep Q-network (DQN) is employed to learn the optimal scheduling policy, which is further refined using policy gradient optimization. This integrated framework aims to maximize task completion efficiency while minimizing energy usage, addressing the complexity of balancing these competing objectives. Extensive simulations validate the proposed approach, demonstrating significant improvements in task completion rates, average travel distances, and energy consumption compared to baseline algorithms such as random scheduling and greedy algorithms. The methodology presents a robust and efficient solution for enhancing operational efficiency in intelligent logistics and warehousing systems.

Keywords: deep reinforcement learning; biomechanical modeling; intelligent logistics; warehousing robots; scheduling optimization; energy efficiency

1. Introduction

The rapid advancement of technology in the logistics and warehousing sector has facilitated the widespread adoption of intelligent robots, aimed at streamlining operations, enhancing efficiency, and reducing costs. However, the effective scheduling of these robots remains a complex and challenging problem, necessitating the development of sophisticated optimization techniques. This study addresses this issue by integrating deep reinforcement learning (DRL) and biomechanical modeling to optimize the scheduling of intelligent logistics and warehousing robots.

Logistics and warehousing operations are critical components of the supply chain, significantly impacting business efficiency and profitability. The integration of robots into these operations has led to notable improvements, such as faster task completion and reduced human labor. Nevertheless, the scheduling of these robots—ensuring optimal task assignment to maximize efficiency and minimize energy consumption—presents a formidable challenge. Traditional scheduling methods, like random and greedy algorithms, often fail to dynamically adapt to changing operational environments and do not consider the physical constraints of the robots.

The scheduling problem's complexity is heightened by the need to balance multiple objectives, including maximizing task completion rates, minimizing travel distances, and reducing energy consumption. These objectives are interdependent and often conflicting, complicating the achievement of a globally optimal solution through conventional approaches.

The importance of optimizing the scheduling of intelligent logistics and warehousing robots cannot be overstated. Efficient scheduling directly enhances operational performance, reduces costs, and boosts overall productivity. Additionally, minimizing energy consumption not only lowers operational expenses but also aligns with global sustainability goals. The necessity of this research is underscored by the limitations of existing scheduling methods, which lack adaptability to dynamic scenarios and fail to incorporate robots' physical constraints, leading to suboptimal performance. The integration of DRL and biomechanical modeling offers a novel approach to address these limitations, leveraging the strengths of both techniques to develop a more robust and efficient scheduling policy.

The primary objective of this study is to develop an optimized scheduling policy for intelligent logistics and warehousing robots by integrating DRL and biomechanical modeling. Specifically, the study aims to: (1) Formulate the scheduling problem as a Markov Decision Process (MDP) to capture the dynamic operational environment; 2) incorporate biomechanical principles to model the physical constraints and energy consumption of the robots; 3) employ a deep Q-network (DQN) to learn the optimal task assignment policy; 4) refine the policy using a policy gradient approach to enhance performance; 5) validate the proposed methodology through extensive simulations and compare its performance against baseline algorithms.

Guiding research questions include: How can the scheduling problem of intelligent logistics and warehousing robots be effectively formulated as an MDP? What are the key biomechanical factors influencing robot energy consumption, and how can they be integrated into the DRL framework? How does the integration of DRL and biomechanical modeling affect the scheduling policy's performance in terms of task completion rates, travel distances, and energy consumption? What are the comparative advantages of the proposed methodology over traditional scheduling algorithms?

To achieve these objectives, the study employs a comprehensive methodology integrating data-driven insights with advanced computational techniques. Data are sourced from a large-scale logistics and warehousing company, providing a detailed sample of robot operations over a six-month period. The methodology involves problem formulation, biomechanical modeling, DRL framework setup, policy gradient optimization, and simulation and validation. The proposed approach is validated through extensive simulations, comparing its performance against baseline algorithms like random and greedy scheduling, using key performance metrics such as task completion rates, average travel distances, and energy consumption.

This study is expected to make significant contributions to the field of intelligent logistics and warehousing by developing a novel and efficient scheduling policy that addresses existing method limitations. The findings aim to enhance operational efficiency, reduce energy consumption, and provide valuable insights for the design and implementation of future intelligent logistics systems.

2. Related works

The integration of deep reinforcement learning (DRL) and biomechanical modeling for optimizing the scheduling of intelligent logistics and warehousing robots is a multidisciplinary field that has garnered significant attention in recent years. Existing research has made substantial progress in both DRL and biomechanical modeling individually, but their combined application for scheduling problems remains an area with potential for further exploration.

In the realm of DRL, several notable works have contributed to the development of algorithms that can handle complex decision-making tasks. Haarnoja et al. [1] introduced the Soft Actor-Critic algorithm, an off-policy actor-critic DRL method that achieves state-of-the-art performance on continuous control benchmark tasks by maximizing entropy, which encourages exploration. This method addresses the challenges of high sample complexity and brittle convergence properties that are common in model-free deep RL algorithms. Similarly, Mnih et al. [2–4] proposed asynchronous methods for deep reinforcement learning, which stabilize training and enable the successful training of neural network controllers on a single multi-core CPU. Their approach demonstrated superior performance on the Atari domain and continuous motor control problems.

Biomechanical modeling, on the other hand, has been applied in various medical and biological contexts. Shao et al. [5] and Shao et al. [6] showcased the application of biomechanical modeling in real-time liver tumor localization using deep learning-assisted methods. These studies demonstrate the accuracy and potential of biomechanical modeling in medical physics, particularly for tracking and localizing tumors with high precision. Furthermore, Van Hasselt et al. [7] and Wang et al. [8] presented advancements in deep reinforcement learning algorithms, such as Double Q-learning and dueling network architectures, which improve value function estimation and policy evaluation, respectively.

Despite these advancements, there is a noticeable gap in the literature regarding the integration of DRL [9] and biomechanical modeling for logistics and warehousing robot scheduling. Existing DRL approaches often focus on simulation environments or simplified real-world scenarios, neglecting the physical constraints and energy consumption patterns of robots [10]. Similarly, while biomechanical models have been successfully applied in medical fields, their application in logistics and warehousing operations is limited [11].

This gap presents an opportunity for the current research, which aims to bridge these two areas by developing a scheduling framework that considers both the decision-making capabilities of DRL and the physical realities imposed by biomechanical modeling [12]. By integrating a deep reinforcement learning framework with a biomechanical model that accurately estimates energy consumption based on robot movements and load, the proposed study seeks to optimize task scheduling for intelligent logistics and warehousing robots [13]. This approach not only addresses the limitations of existing DRL methods in real-world applications but also leverages the precision of biomechanical modeling to enhance the efficiency and sustainability of robot operations [14].

The current research will contribute to the field by providing a comprehensive analysis of the scheduling problem, taking into account the complexities of both DRL and biomechanical modeling [15]. Through the development and validation of an integrated framework, this study will offer a novel perspective on optimizing robot scheduling, potentially leading to more efficient and cost-effective logistics and warehousing operations [16,17].

The combination of DRL and biomechanical modeling brings unique advantages to many fields, such as medicine, robotics, sports science and so on. This cross-integration not only improves the authenticity and practicability of the model, but also provides a new way to solve complex problems.

1) Improve model performance

More accurate motion simulation: Biomechanical modeling can accurately describe the physical structure and motion principles of organisms, while deep reinforcement learning can optimize motion strategies by learning a large amount of data. The combination of the two can make the simulated biological movement closer to the real situation. When simulating human walking, running and other complex movements, it can more accurately capture the changes of muscle strength, joint angle and other factors, and provide a more reliable reference for sports training, rehabilitation treatment and other fields.

Enhanced environmental adaptability: Deep reinforcement learning has a powerful learning ability, allowing models to learn and adapt under different environmental conditions. Combined with biomechanical modeling, the model can not only take into account the mechanical characteristics of the organism itself but also adjust the movement strategy in real time according to the changes of the environment. For example, in the task of robot navigation, the model combined with biomechanics can adjust the way of walking or moving according to factors such as the fluctuation of terrain and the change of friction so as to improve the adaptability and stability of robots in complex environments.

2) Excavate the biomechanical mechanism

Revealing the principle of motion control: Deep reinforcement learning can discover hidden patterns and laws from a large amount of biological motion data. Combined with biomechanical modeling, it helps to reveal how organisms control muscles and bones through the nervous system to achieve various movements. By analyzing the movement strategies learned by the model, we can deeply understand the control mechanism of organisms in different movement tasks, and provide new perspectives and methods for the study of neuroscience and exercise physiology.

Exploring biological evolution strategies: The application of deep reinforcement learning to biomechanical modeling can simulate the motor adaptation process of organisms during long-term evolution. By setting different environmental pressures and survival goals, the model can learn the movement strategies similar to biological evolution, and help us understand how organisms adapt to environmental changes by constantly adjusting their mechanical structure and movement mode, thus promoting the development of biological evolution theory.

3) Optimize engineering design

Design of power-assisted robot: In the design of the robot, it is an important way to improve the performance of the robot by using the mechanical structure and motion mode of biology for reference. The combination of deep reinforcement learning and biomechanical modeling can help engineers better understand the advantages of biological motion and apply these advantages to the structural design and control algorithms of robots. The movement of the robot is more efficient, flexible and stable, and the performance of the robot in various tasks is improved.

Improve the design of rehabilitation equipment: For the design of rehabilitation equipment, combined with deep reinforcement learning and biomechanical modeling, personalized rehabilitation training programs can be customized according to the specific conditions of patients, such as muscle strength, joint range of motion, etc. At the same time, it can also optimize the mechanical properties of rehabilitation equipment, so that it can better assist patients in rehabilitation training and improve the rehabilitation effect.

Amazon's e-commerce warehouse uses traditional DRL algorithms (such as DQN) to optimize the path planning of AGV (automatic guided vehicle), but when orders surge during peak periods, the frequent start and stop of the robot leads to overheating of the motor and a sharp drop in battery life. DRL improves sorting efficiency by 20% after implementing the new technology, but increases hardware attrition by 30%. The main problem is that the traditional DRL does not consider the kinematic constraints of the robot (such as acceleration inertia, joint torque constraints), which leads to the mismatch between action instructions and physical execution.

The main solution is to introduce the energy consumption model in biped robot gait optimization, combine the joint torque, motor power and DRL reward function, and constrain the acceleration and steering frequency of the robot. The results show that the energy consumption is reduced by 18%, the hardware failure rate is reduced by 25%, and the sorting efficiency is only sacrificed by 5%. A multi-objective DRL framework is proposed, in which energy consumption, efficiency and mechanical loss are included in the optimization objective at the same time, rather than a single sorting efficiency index.

3. Method

3.1. Data source

The data utilized in this study were sourced from a large-scale logistics and warehousing company, providing detailed records of robot operations, including task assignments, travel times, and energy consumption. The dataset spans a period of six months, ensuring a comprehensive and representative sample of operational scenarios. To protect proprietary information and comply with privacy regulations, the data were anonymized.

Table 1. Sample dataset structure.

Robot ID	Task type	Start time	End time	Travel distance (m)	Energy consumption (kWh)
001	Pickup	08:00	08:15	120	0.5
002	Delivery	08:05	08:20	150	0.6
003	Restock	08:10	08:30	200	0.7
004	Pickup	08:15	08:25	100	0.4
005	Delivery	08:20	08:35	180	0.65

Table 1 illustrates the structure and content of the dataset, presenting key

variables such as robot ID, task type, start time, end time, travel distance, and energy consumption.

3.2. Research methodology

The integration of deep reinforcement learning (DRL) and biomechanical modeling to optimize the scheduling of intelligent logistics and warehousing robots involves several key steps [18]. The methodology is structured as follows:

1) Problem formulation: The scheduling problem is formulated as a Markov Decision Process (MDP) [19]. The state space $\mathcal{S}$ includes the current positions of all robots, task statuses, and robot energy levels. The action space $\mathcal{A}$ consists of possible task assignments for each robot. The reward function $R(s,a)$ is designed to maximize task completion efficiency while minimizing energy consumption:

$$R(s,a) = \alpha \cdot \frac{\text{Number of completed tasks}}{\text{Total tasks}} - \beta \cdot \text{Energy consumed}.$$

where α and β are weighting parameters.

2) Biomechanical modeling: To accurately model the physical constraints and energy consumption of the robots, biomechanical principles are incorporated [20–23]. The energy consumption E for a robot traveling a distance d is estimated using:

$$E = \gamma \cdot d + \delta \cdot d \cdot \text{Load}.$$

where γ and δ are constants derived from biomechanical experiments, and Load represents the weight of the items being transported.

3) Deep reinforcement learning framework: A deep Q-network (DQN) is employed to learn the optimal policy $\pi^*(s)$ [24,25]. The Q-value function $Q(s,a)$ is approximated using a neural network with parameters θ:

$$Q(s,a;\theta) \approx Q^*(s,a).$$

The loss function for the DQN is defined as:

$$L(\theta) = \mathbb{E}[(Q(s,a;\theta) - y)^2].$$

where $y = R(s,a) + \gamma \max_{a'} Q(s',a';\theta^-)$, and θ^- represents the target network parameters.

4) Policy gradient optimization: To refine the policy, a policy gradient approach is adopted. The gradient of the expected reward with respect to the policy parameters ϕ is computed as:

$$\nabla_\phi J(\pi_\phi) = \mathbb{E}_{\pi_\phi}\left[\nabla_\phi \log \pi_\phi(a|s) Q(s,a)\right].$$

The policy parameters are updated using stochastic gradient ascent:

$$\phi \leftarrow \phi + \alpha \nabla_\phi J(\pi_\phi).$$

5) Integration of DRL and biomechanical model: The biomechanical model is integrated into the DRL framework by incorporating energy consumption estimates into the reward function [26]. This ensures the learned policy optimizes both task completion and physical constraints.

6) Simulation and validation: The proposed methodology is validated through extensive simulations using the collected dataset. The performance of the optimized scheduling policy is compared against baseline algorithms, such as random scheduling and greedy algorithms.

The application effects of deep reinforcement learning (DRL) and biomechanical modeling in optimizing the scheduling of intelligent logistics and warehouse robots can be evaluated from the dimensions of task execution, system performance, cost-effectiveness, and technological innovation, as follows:

1) Task execution effects

Task completion rate: It measures the ratio of the number of tasks successfully completed by the robots within a given time to the total number of tasks. If the task completion rate significantly improves after combining DRL and biomechanical modeling, approaching or reaching 100%, it indicates that the robots can be effectively scheduled to complete various logistics and warehouse tasks.

Task execution accuracy: It assesses the accuracy of the robot's operations during task execution, such as cargo handling and storage location placement. It can be measured by calculating the number of operation errors or the error range. A high standard of accuracy, such as the cargo placement error within the prescribed millimeter range, indicates a good optimization effect.

Complex task handling capability: Observe the robot's performance when dealing with complex tasks such as simultaneous handling of multiple cargos and conflicts in path planning. If it can quickly and reasonably plan paths and allocate tasks without long-term stagnation or confusion, it indicates that the optimization helps enhance the robot's ability to cope with complex scenarios.

2) System performance indicators

Robot operation efficiency: Including the walking speed of the robots, the speed of cargo loading and unloading, etc. The average time for the robot to complete a single task or tasks within a unit time before and after optimization can be compared. If the average task time is shortened, it indicates an improvement in operation efficiency.

Resource utilization rate: Analyze the proportion of the robot's working time to the total time, energy consumption, etc. If the resource utilization rate increases, such as an increase in the proportion of the robot's working time and reasonable energy consumption, it indicates that the scheduling optimization enables the robot to utilize resources more fully.

System stability: Statistics on the frequency of failures, freezes, or abnormalities during the system's operation. A high stability means that the system can operate stably for a long time, and there are few problems such as robot collisions and task interruptions caused by unreasonable scheduling.

3) Cost-effectiveness assessment

Operating cost: Calculate the total of human costs, equipment maintenance costs, energy costs, etc., of the logistics and warehouse system after optimization using DRL and biomechanical modeling. If the operating cost reduces, such as reducing the number of robots or reducing energy consumption through optimized scheduling, it indicates that the optimization is cost-effective.

Economic benefit: Compare the economic benefits brought by business indicators such as throughput and order processing volume of logistics and warehouse business before and after optimization. If the economic benefit significantly improves, such as an increase in order processing volume and an acceleration of cargo turnover speed bringing more income, it indicates that the optimization application effect is good.

3.3. Mathematical formulations

The following mathematical formulations detail the key components of the methodology:

1) State representation:

$$s = (p_1, p_2, \dots, p_N, t_1, t_2, \dots, t_M, e_1, e_2, \dots, e_N).$$

where p_i represents the position of robot i, t_j represents the status of task j, and e_i represents the energy level of robot i.

2) Action representation:

$$a = (a_1, a_2, \dots, a_N).$$

where a_i represents the task assigned to robot i.

3) Transition function:

$$s' = T(s, a).$$

where T is the transition function mapping the current state and action to the next state.

4) Reward function:

$$R(s, a) = \alpha \cdot \frac{\sum_{j=1}^{M} \text{completed}(t_j)}{M} - \beta \cdot \sum_{i=1}^{N} E_i.$$

where $\text{completed}(t_j)$ is an indicator function returning 1 if task j is completed.

5) Energy consumption model:

$$E_i = \gamma \cdot d_i + \delta \cdot d_i \cdot \text{Load}_i.$$

6) Q-value update:

$$Q(s, a; \theta) \leftarrow Q(s, a; \theta) + \alpha[R(s, a) + \gamma \max_{a'} Q(s', a'; \theta^-) - Q(s, a; \theta)].$$

7) Policy gradient update:

$$\phi \leftarrow \phi + \alpha \sum_{a \in \mathcal{A}} \pi_\phi(a|s) \nabla_\phi \log \pi_\phi(a|s) Q(s, a).$$

8) Expected reward:

$$J(\pi_\phi) = \mathbb{E}_{\pi_\phi}[R(s, a)].$$

4. Results

4.1. Performance metrics

To evaluate the effectiveness of the proposed methodology, several key performance metrics were monitored, including task completion rate, average travel distance, and energy consumption. The results are compared against baseline algorithms such as random scheduling and greedy algorithms.

As shown in **Table 2**, the proposed method significantly outperforms the baseline algorithms in terms of task completion rate.

Table 2. Comparison of task completion rates.

Algorithm	Task completion rate (%)
Random scheduling	75.2
Greedy algorithm	82.5
Genetic algorithm	75.4
Particle swarm optimization	81.3
Hierarchic genetic algorithm	85.3
Proposed method	90.3

Table 3 illustrates that the proposed method reduces the average travel distance per task compared to the baseline algorithms.

Table 3. Average travel distance per task.

Algorithm	Average travel distance (m)
Random scheduling	175.4
Greedy algorithm	160.2
Genetic algorithm	187.8
Particle swarm optimization	178.4
Hierarchic genetic algorithm	158.7
Proposed method	145.8

As depicted in **Table 4**, the proposed method achieves lower energy consumption per task than the baseline algorithms.

Table 4. Average energy consumption per task.

Algorithm	Average energy consumption (kWh)
Random scheduling	0.85
Greedy algorithm	0.78
Genetic algorithm	0.86
Particle swarm optimization	0.76
Hierarchic genetic algorithm	0.72
Proposed method	0.65

Table 5 shows that different algorithms evaluate the response time of the running process, and the method used in this paper has a good application in response time.

Table 5. Comparison of response time for different models.

Algorithm	Response time (ms)
Random scheduling	64.6
Greedy algorithm	48.1
Genetic algorithm	53.7
Particle swarm optimization	46.5
Hierarchic genetic algorithm	43.7
Proposed method	26.4

4.2. Multi-objective tradeoff experimental design

The reward function variables are defined as follows:

$$R = \alpha \cdot \text{Refficiency} + \beta \cdot \text{Renergy} + \gamma \cdot \text{Rsafety}, \text{ Restraint: } \alpha + \beta + \gamma = 1.$$

The balance relationship between the two is shown in **Table 5** below, where $\gamma = 0.1$ is fixed for a typical scene, and the balance between α and β is focused.

Table 6. The balance relationship between α and β.

Scene type	Characterization	Experimental parameter range
Peak period	Order volume surges by 300%, priority given to timeliness	$\alpha \in [0.6, 0.8]$ $\beta \in [0.2, 0.4]$
Steady period	Equilibrium mode	$\alpha = \beta = 0.45$
Equipment maintenance period	Need to extend hardware lifespan	$\beta \in [0.7, 0.9]$
Green storage policy	Hard constraints on government energy consumption indicators	$\beta \geq 0.6$

Efficiency sensitive area ($\alpha > 0.7$): The task time is shortened by 12%, but the energy consumption is increased by 25%, which is suitable for the high timeliness scenario (such as luxury warehouse) with a unit price > 500 yuan.

Energy consumption sensitive area ($\beta > 0.6$): The energy consumption is reduced by 18%, and the task time is extended by 8%, which is suitable for non-emergency operations such as night replenishment.

Equilibrium zone ($\alpha = 0.5$, $\beta = 0.4$): Within ±5% change in time/energy consumption, suitable for daily operation.

4.3. Detailed results

The detailed results further break down the performance metrics across different operational scenarios, highlighting the consistency and robustness of the proposed method.

Table 7 shows the task completion rates across various demand scenarios, confirming the superior performance of the proposed method.

Table 7. Task completion rate across different scenarios.

Scenario	Low demand (%)	Medium demand (%)	High demand (%)
Random scheduling	78.5	74	72.3
Greedy algorithm	85.2	81.5	79.8
Genetic algorithm	75.9	73.5	71.7
Particle swarm optimization	83.4	82.6	80.2
Hierarchic genetic algorithm	85.6	84.5	82.4
Proposed method	92.1	89.6	88

Table 8 demonstrates the average travel distances across different scenarios, with the proposed method consistently showing lower values.

Table 8. Average travel distance across different scenarios.

Scenario	Low demand (m)	Medium demand (m)	High demand (m)
Random scheduling	170.2	178.6	180.5
Greedy algorithm	155.4	163.2	165.8
Genetic algorithm	172.5	175.6	181.6
Particle swarm optimization	168.5	172.4	176.8
Hierarchic genetic algorithm	150.3	156.3	162.4
Proposed method	140.1	148.5	150.3

Table 9 presents the average energy consumption across various demand scenarios, further validating the energy efficiency of the proposed method.

Table 9. Average energy consumption across different scenarios.

Scenario	Low demand (m)	Medium demand (m)	High demand (m)
Random scheduling	0.82	0.87	0.88
Greedy algorithm	0.76	0.8	0.81
Genetic algorithm	0.78	0.81	0.84
Particle swarm optimization	0.72	0.76	0.79
Hierarchic genetic algorithm	0.69	0.72	0.76
Proposed method	0.63	0.67	0.68

These tables collectively demonstrate the superior performance of the proposed integration of DRL and biomechanical modeling in optimizing the scheduling of intelligent logistics and warehousing robots. The results indicate significant improvements in task completion rates, reductions in travel distances, and lower energy consumption compared to traditional scheduling algorithms.

5. Discussion

5.1. Significance of results

The integration of deep reinforcement learning (DRL) and biomechanical modeling for optimizing the scheduling of intelligent logistics and warehousing robots has yielded several pivotal insights and demonstrated notable enhancements in operational efficiency and energy consumption. This section elucidates the significance of the obtained results, emphasizing the method's efficacy in addressing complex scheduling challenges.

As illustrated in **Table 2**, the task completion rate improved from 75.2% under random scheduling and 82.5% with the greedy algorithm to 90.3% using the proposed method. This substantial increase underscores the capability of the DRL-based approach, augmented by biomechanical modeling, to more effectively allocate tasks, thereby maximizing operational throughput.

Additionally, the reduction in average travel distance per task, as shown in **Table 3**, decreased from 175.4 m (random scheduling) and 160.2 m (greedy algorithm) to 145.8 m (proposed method). This optimization in path planning not only expedites task completion but also minimizes wear and tear on the robots, potentially extending their operational lifespan.

A particularly compelling outcome is the reduction in energy consumption, detailed in **Table 4**, which decreased from 0.85 kWh (random scheduling) and 0.78 kWh (greedy algorithm) to 0.65 kWh (proposed method). This significant energy saving highlights the importance of incorporating biomechanical principles into the DRL framework, ensuring that robots operate efficiently within their physical constraints.

The consistency of these improvements across various operational scenarios further validates the robustness of the proposed method. Whether under low, medium, or high demand, the integrated approach consistently outperformed traditional algorithms, demonstrating its adaptability and reliability in diverse operational conditions.

5.2. Innovative aspects

The innovation in this study lies in the seamless integration of DRL and biomechanical modeling. Traditional DRL approaches often overlook the physical limitations of agents, focusing solely on decision-making optimization. By incorporating biomechanical principles, this study bridges a critical gap, ensuring that the learned policies are both optimal in task allocation and feasible given the physical constraints of the robots.

The utilization of a Markov Decision Process (MDP) to formulate the scheduling problem provides a structured framework for applying DRL. The inclusion of energy consumption in the reward function is a novel approach that aligns optimization objectives with real-world operational constraints. This dual-objective optimization, which maximizes task completion while minimizing energy consumption, represents a significant advancement over single-objective methods.

Furthermore, the policy gradient optimization step refines the learned policies, enhancing their robustness and adaptability. This multi-faceted approach ensures that the scheduling policy is not only efficient but also responsive to changing operational conditions.

The corresponding implementation process is given for warehousing enterprises of different scales:

1) The implementation steps of large enterprises are mainly as follows:

- Data layer:

Multi-modal data acquisition: mechanical sensors (sampling rate ≥ 1 kHz) and environmental monitoring modules (temperature, humidity and illumination) are deployed at AGV joints, shelves and charging piles.

Annotation rules: associate hardware loss events (such as motor overheating code E103) with action sequences to build a biomechanics-control combined tag library.

- Model layer:

Parameter adjustment: Pre-train the biomechanical sub-model based on the historical data of the enterprise, fix the joint inertia parameters (such as moment of inertia), and only fine-tune the reward weight of DRL. Enable distributed DRL training, divide agents according to warehouse partition, and learn shared policies through Federation.

- System integration:

API middleware: The DRL decision engine is encapsulated as a RESTful service, which interacts with the order pool and path planning module of the existing WMS (warehouse management system) through Kafka.

Hardware adaptation: deploy a lightweight biomechanical model (TensorRT acceleration, model volume < 20 MB) in the robot embedded system, and feed back the joint torque to the DRL decider in real time.

2) The implementation steps for SMEs are as follows:

- Lightweight deployment:

Model compression: Knowledge distillation technology is used to migrate large enterprise pre-trained models to lightweight architectures (such as MobileNet V3 + PPO), and the accuracy loss is controlled within 5%.

Parameter adjustment: Turn off complex environmental response modules (such as dynamic shock absorbers) and focus on core indicators.

- Progressive integration:

Hybrid Scheduling: In existing rule-based scheduling systems such as Lefthand Robotics, the DRL optimization module is enabled only for peak hour orders.

Low-cost sensing: Use an RGB-D camera instead of a high-precision force sensor to estimate the load moment by visually extrapolating the shelf deformation (error < 8%).

- Cloud edge collaboration:

Training on the cloud: Use AWS RoboMaker or Alibaba Cloud intelligent computing to complete model training, reducing the cost by 60%.

Inference on the side: performing real-time decisions locally via NVIDIA Jetson AGX Xavier.

5.3. Limitations

Despite the promising outcomes, several limitations warrant consideration. Firstly, the dataset, although comprehensive, originates from a single logistics and warehousing company, potentially limiting the generalizability of the findings to other contexts with distinct operational dynamics and robot specifications.

Secondly, the biomechanical model employed in this study is based on specific assumptions and constants derived from experiments. Variations in robot design, load characteristics, and environmental conditions could impact the accuracy of energy consumption estimates. Future research should explore more adaptive biomechanical models capable of adjusting to different operational contexts.

Thirdly, the computational complexity of the DRL framework, particularly in large-scale operations with numerous robots and tasks, poses practical challenges. The scalability of the proposed method needs further investigation to ensure its feasibility in real-world applications.

Lastly, the study assumes a static environment with known task locations and robot positions. Dynamic environments with unpredictable changes could affect the performance of the scheduling policy. Incorporating real-time adaptability into the DRL framework could mitigate this limitation.

In summary, while the integration of DRL and biomechanical modeling provides a robust solution to the scheduling problem in intelligent logistics and warehousing, addressing these limitations is essential for realizing its full potential in diverse and dynamic operational settings. Future research should focus on enhancing the generalizability, adaptability, and scalability of the proposed methodology.

6. Conclusion

6.1. Summary

This study integrates deep reinforcement learning (DRL) and biomechanical modeling to address the scheduling challenges of intelligent logistics and warehousing robots, utilizing a comprehensive dataset from a large-scale logistics company. The primary findings indicate that the proposed methodology substantially enhances operational efficiency and reduces energy consumption compared to traditional scheduling algorithms.

6.2. Key findings

1) Improved task completion rates: The proposed method achieved a task completion rate of 90.3%, significantly outperforming random scheduling (75.2%) and the greedy algorithm (82.5%).
2) Reduced travel distances: The average travel distance per task was reduced to 145.8 m, marking a notable decrease from 175.4 m (random scheduling) and 160.2 m (greedy algorithm).
3) Lower energy consumption: The average energy consumption per task was minimized to 0.65 kWh, compared to 0.85 kWh (random scheduling) and 0.78 kWh (greedy algorithm).

6.3. Contributions to the field

This research makes several contributions to the field of intelligent logistics and warehousing:

- Innovative integration: It pioneers the integration of DRL with biomechanical modeling, offering a novel approach to optimize robot scheduling that accounts for both operational efficiency and physical constraints.

- Enhanced performance metrics: The study demonstrates significant improvements in key performance metrics, providing a robust solution for real-world logistics challenges.
- Data-driven validation: The use of a large, real-world dataset ensures the validity and applicability of the findings, bridging the gap between theoretical research and practical implementation.

6.4. Practical applications and recommendations

The findings of this research hold substantial practical value for logistics and warehousing operations:

- Operational efficiency: The optimized scheduling policy can be directly applied to enhance the efficiency of robot operations, leading to faster task completion and higher throughput.
- Energy savings: By reducing energy consumption, the proposed method contributes to cost savings and promotes sustainability in logistics operations.
- Scalability: The methodology is scalable and adaptable to various operational scales and scenarios, making it versatile for different warehouse environments.

Recommendations for practice:

- Implementation strategy: Companies should consider a phased implementation of the proposed scheduling policy, initiating with a pilot program to fine-tune parameters and ensure seamless integration with existing systems.
- Continuous monitoring and adaptation: Regular monitoring of performance metrics and adaptive adjustments to the model are crucial to maintaining optimal performance in dynamic operational environments.
- Training and support: Providing training for operational staff and technical support during the integration process will facilitate smoother adoption and maximize the benefits of the optimized scheduling system.

In conclusion, the integration of DRL and biomechanical modeling offers a transformative approach to optimizing the scheduling of intelligent logistics and warehousing robots, delivering significant improvements in efficiency and energy consumption. The practical applications and recommendations outlined provide a roadmap for organizations to harness these advancements in their operations.

Conflict of interest: The author declares no conflict of interest.

References

1. Haarnoja T, Zhou A, Abbeel P, Levine S. Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor. arXiv. 2018; arXiv:1801.01290.
2. Mnih V, Badia AP, Mirza M, et al. Asynchronous Methods for Deep Reinforcement Learning. arXiv. 2016; arXiv:1602.01783.
3. Mnih V, Kavukcuoglu K, Silver D, et al. Playing Atari with Deep Reinforcement Learning. arXiv. 2013; arXiv:1312.5602.
4. Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning. Nature. 2015; 518(7540): 529-533. doi: 10.1038/nature14236
5. Shao HC, Wang J, Bai T, et al. Real-time liver tumor localization via a single x-ray projection using deep graph neural network-assisted biomechanical modeling. Physics in Medicine & Biology. 2022; 67(11): 115009. doi: 10.1088/1361-6560/ac6b7b

6. Shao H, Huang X, Folkert MR, et al. Automatic liver tumor localization using deep learning-based liver boundary motion estimation and biomechanical modeling (DL-Bio). Medical Physics. 2021; 48(12): 7790-7805. doi: 10.1002/mp.15275
7. Van Hasselt H, Guez A, Silver D. Deep Reinforcement Learning with Double Q-Learning. Proceedings of the AAAI Conference on Artificial Intelligence. 2016; 30(1). doi: 10.1609/aaai.v30i1.10295
8. Wang Z, Schaul T, Hessel M, et al. Dueling Network Architectures for Deep Reinforcement Learning. arXiv. 2015; arXiv:1511.06581.
9. Lillicrap TP, Hunt JJ, Pritzel A, et al. Continuous control with deep reinforcement learning. arXiv. 2015; arXiv: 1509.02971.
10. Hessel M, Modayil J, Van Hasselt H, et al. Rainbow: Combining Improvements in Deep Reinforcement Learning. Proceedings of the AAAI Conference on Artificial Intelligence. 2018; 32(1). doi: 10.1609/aaai.v32i1.11796
11. Vogt N. Biomechanical modeling of whole bodies. Nature Methods. 2024; 21(12): 2228-2228. doi: 10.1038/s41592-024-02548-4
12. Lindquist Liljeqvist M, Bogdanovic M, Siika A, et al. Geometric and biomechanical modeling aided by machine learning improves the prediction of growth and rupture of small abdominal aortic aneurysms. Scientific Reports. 2021; 11(1). doi: 10.1038/s41598-021-96512-3
13. Ruijsink B, Zugaj K, Wong J, et al. Dobutamine stress testing in patients with Fontan circulation augmented by biomechanical modeling. PLOS ONE. 2020; 15(2): e0229015. doi: 10.1371/journal.pone.0229015
14. Lee CH, Laurence DW, Ross CJ, et al. Mechanics of the Tricuspid Valve—From Clinical Diagnosis/Treatment, In-Vivo and In-Vitro Investigations, to Patient-Specific Biomechanical Modeling. Bioengineering. 2019; 6(2): 47. doi: 10.3390/bioengineering6020047
15. Zhang Y, Huang X, Wang J. Advanced 4-dimensional cone-beam computed tomography reconstruction by combining motion estimation, motion-compensated reconstruction, biomechanical modeling and deep learning. Visual Computing for Industry, Biomedicine, and Art. 2019; 2(1). doi: 10.1186/s42492-019-0033-6
16. Miller K, Joldes GR, Bourantas G, et al. Biomechanical modeling and computer simulation of the brain during neurosurgery. International Journal for Numerical Methods in Biomedical Engineering. 2019; 35(10). doi: 10.1002/cnm.3250
17. Chanda A. Biomechanical Modeling of Human Skin Tissue Surrogates. Biomimetics. 2018; 3(3): 18. doi: 10.3390/biomimetics3030018
18. Zhou M. Analysis Of Current Development Status of Global Path Planning for Intelligent Carrier Robots in Warehouse Logistics Scenarios. Highlights in Science, Engineering and Technology. 2024; 103: 14-19. doi: 10.54097/3zcsde07
19. Yang M, Sun J, Li J. Research on Hot Spots and Development Trends of Intelligent Logistics in the Era of Artificial Intelligence an Analysis of Knowledge Graph Based on CiteSpace. In: Proceedings of the 4th International Conference on Artificial Intelligence and Computer Engineering. 2023. pp. 586-591. doi: 10.1145/3652628.3652726
20. Duan LM. Path Planning for Batch Picking of Warehousing and Logistics Robots Based on Modified A* Algorithm. International Journal of Online and Biomedical Engineering (iJOE). 2018; 14(11): 176-192. doi: 10.3991/ijoe.v14i11.9527
21. Tian R, Zhu L, Cai Z. Research on task scheduling and path planning in intelligent warehousing system. In: Proceedings of the Third International Conference on Control and Intelligent Robotics (ICCIR 2023). p. 127. doi: 10.1117/12.3011342
22. Mo Y, Sun Z, Yu C. EventTube: An Artificial Intelligent Edge Computing Based Event Aware System to Collaborate With Individual Devices in Logistics Systems. IEEE Transactions on Industrial Informatics. 2023; 19(2): 1823-1832. doi: 10.1109/tii.2022.3189177
23. Han J, Zhao J, Li R. Dynamic Path Planning of Intelligent Warehouse Robot Based on Improved A* Algorithm. Computer-Aided Design and Applications. 2024; 134-143. doi: 10.14733/cadaps.2024.s23.134-143
24. Gui Y. Logistics Robot Path Planning: A* Algorithm Operation and Analysis. Journal of Physics: Conference Series. 2023; 2547(1): 012028. doi: 10.1088/1742-6596/2547/1/012028
25. Yuan R, Pan L, Li J, et al. Road Network Optimization of Intelligent Warehouse Picking Systems Based on Improved Genetic Algorithm. In: Proceedings of the 2021 IEEE 7th International Conference on Cloud Computing and Intelligent Systems (CCIS). 2021. pp. 361-366. doi: 10.1109/ccis53392.2021.9754603
26. Liu Y, Li J. Runtime Verification-Based Safe MARL for Optimized Safety Policy Generation for Multi-Robot Systems. Big Data and Cognitive Computing. 2024; 8(5): 49. doi: 10.3390/bdcc8050049