Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 1, pp. 241–249, DOI 10.12694/scpe.v26i1.3765

THE FACTORY SUPPLY CHAIN MANAGEMENT OPTIMIZATION MODEL BASED ON DIGITAL TWINS AND REINFORCEMENT LEARNING

XINBO ZHAO* AND ZHIHONG WANG†

Abstract. This paper introduces the "digital twin" to solve the problem of material allocation and real-time scheduling in the warehouse site. This project intends first to establish mathematical modeling based on a digital twin unmanned warehouse and dynamically optimize materials in the unmanned warehouse by combining visual analysis and deep reinforcement learning. Then, a security sharing mechanism of digital twin-edge network data based on blockchain fragmentation is proposed. For twin models with time-varying characteristics, a multi-node adaptive resource optimization method such as multipoint cluster selection, local base station consistent access selection, spectrum and computational consistency is constructed. This is done to maximize blockchain business processing power. A two-layer near-end strategy optimization (PPO) algorithm is proposed to solve the adaptive resource optimization problem. Experiments have proved that this method can significantly improve the overall processing power of the blockchain. In addition, this method is more adaptable than conventional deep reinforcement learning.

Key words: Unmanned storage; Digital twins; Deep reinforcement learning; Dynamic scheduling optimization; Digital twin edge network; Blockchain sharding

1. Introduction. To achieve accurate scheduling and optimal allocation of resources, most of the existing methods use heuristic methods to convert multiple high-quality multi-objective programming problems into a single programming problem. Alternatively, vehicles, three-dimensional shelves, testing equipment, etc., are regarded as a resource, and the optimal decision method is adopted to solve the problem [1]. However, the existing methods are limited in computing power and flexibility and can not effectively deal with multi-frequency, uncertain quantity of goods arrival, shelf, AGV, forklift and other resource optimization allocation problems. This seriously affects the service level and efficiency of the warehouse system. As a frontier and hot spot in intelligent manufacturing and storage, a digital twin is introduced into unmanned storage in this paper. Literature [2] takes the digital twin five-dimensional model as an example to introduce the application of this model in the warehouse. However, this method is mainly used in the manufacturing industry and can only play a reference role. Literature [3] uses digital twins to develop a new multi-mode intelligent terminal to solve the problem that real-time interaction cannot occur in manufacturing. Literature [4] integrates cyber twins with digital twins to build a networked digital twin model and remote control system oriented to information-physical fusion. This paper takes "digital twin" and "unmanned storage" as the starting point to study the integration of "multi-class resource scheduling" and "efficient scheduling." The working condition of the equipment is monitored in real-time utilizing the Internet and visualization to improve its working efficiency and accurate scheduling level. Therefore, the digital twin unmanned warehouse architecture with multi-level characteristics is constructed according to the characteristics of the unmanned warehouse operation process. A real-time map construction method of unmanned warehouses based on physical modeling and data service systems is proposed. Then, the resource scheduling problem of a digital twin unmanned warehouse based on deep reinforcement learning is studied utilizing multidimensional information fusion.

2. Digital twin unmanned storage system design.

2.1. System Architecture. This paper establishes the architecture of a digital Twin unmanned warehouse (Figure 2.1 is referenced in Building the Digital Representation with Digital Twin using Microsoft stack).

*Liaoning University of International Business and Economics, Dalian Liaoning 116052, China
†Liaoning University of International Business and Economics, Dalian Liaoning 116052, China (Corresponding author, zhaoxinbo0124@ 163.com)

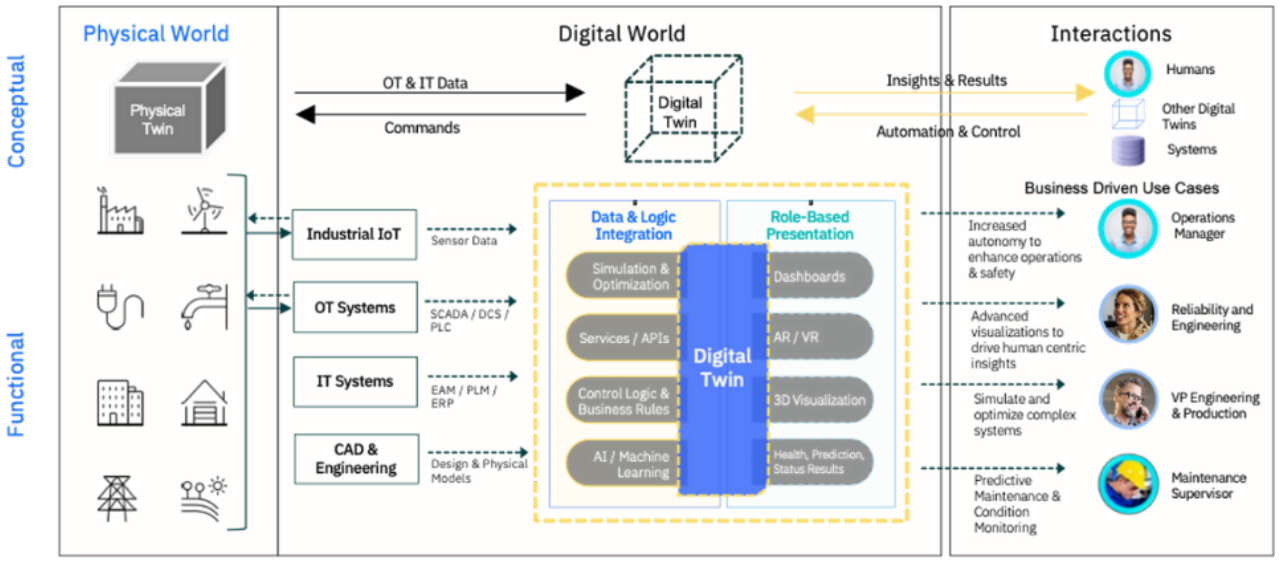

Fig. 2.1: Architecture of digital twin unmanned storage system.

The model consists of a visual, physical entity layer, two technology platforms, three layers, three databases, six logical process mappings, and six physical realities time mapping.

Among them, the sensing layer mainly identifies and acquires the target [5]. This layer uses the sensor node method to process the relevant data of shelves, forklifts, AGVs, goods, pallets, robots, warehouses, etc. and then transmits it to the corresponding location through relevant networking means to realize the collection of information on the lower layer.

At the data level, it realizes the management of user rights, the model interaction interface between the target model base, the real-time database and the local database [6]. Including warehouse signal, equipment status, equipment location, display information, warehouse location information, etc. Local databases include layout data, logical data, trigger mapping, initialization rules, scan point mapping tables, configuration files, etc. The system includes equipment data, production data, model base, operation base, order information, user personal information and so on.

At the business level, a predictive data-driven modeling method is adopted based on the twin data. This will make the warehouse management intelligent, thereby optimizing resource efficiency, optimizing the number of orders, optimizing the warehouse location, optimizing the area, and sharing resource information [7]. At the same time, the optimal results will also be transmitted back to the data center of the perception layer for virtual monitoring of the perception layer.

2.2. Elements of the operation process of digital twin unmanned warehouses. Among them, the operation process of a digital twin unmanned warehouse includes establishing the twin entity model, data system, and mapping logic.

2.2.1. Twin entity modeling. The ontology modeling method is used first when constructing twin entities. The ontology is constructed with class and attribute as the core. The category refers to the definition of the entity, and the property is the expression of the specific role of the class. The target and its properties must be modified before creation, and then its output is stored in the object library and recorded in the target table [8]. Lightweight methods are selectively adopted to reduce the display load during operation. The movable part in the 3D model is set as a movable body, and then the behavior trajectory of the movable part is modified. Animate it with associated components to form a whole. The elements of an unmanned warehouse and their relationship together constitute a complex network concept system. It includes 16 categories of objects, 21 connections, and 91 properties. You can see a detailed description in Figure 2.2.

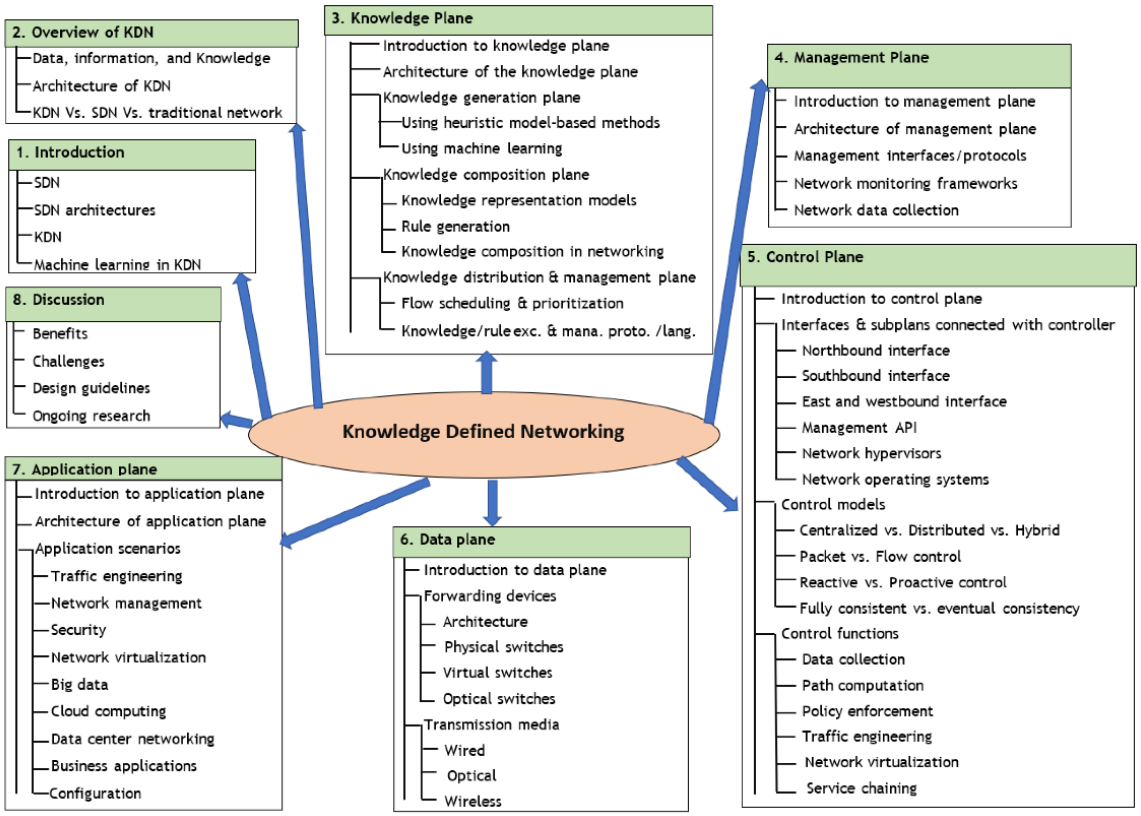

Fig. 2.2: Network structure of unmanned storage ontology model.

2.2.2. Data System Construction. Data service systems can realize real-time connections and calls between local and system databases. Run with a real-time database-driven model. A dynamic database-driven model is adopted. Entity knowledge ontology based on the OWL method is used to realize dynamic access to dynamic data in a database [9]. The database interface module is accessed regularly to realize the data analysis of the local database and system database. In this way, the unmanned storage environment can be quickly restored.

2.2.3. Mapping Logic. A geometric modeling method based on ontology is proposed, which can realize the unity of objects in space position, geometric size, motion characteristics, etc. The data service platform provides a unified control interface inside and outside the schema and interacts with the three databases [10]. Based on the law of real-time mapping, this paper efficiently combines the physical elements of the digital twin model to run the process of warehousing, tallying, storage, picking, order receiving, and delivery to the online process. This forms a complete unmanned warehouse business process. The logical flow of the real-time mapping is shown in Figure 2.3 (image cited in Developments in the Built Environment, Volume 17, March 2024, 100309).

2.3. Digital twin unattended warehouse scheduling optimization logic. The digital twin method is used to realize the effective utilization of the warehouse. Cluster analysis and deep reinforcement learning are used to analyze and optimize the resource effectiveness of the system [11]. After resource efficiency optimization, the allocation scheme is compared with that before optimization and fed back to the database for vector iteration to get the optimal solution. The specific content is shown in Figure 2.4.

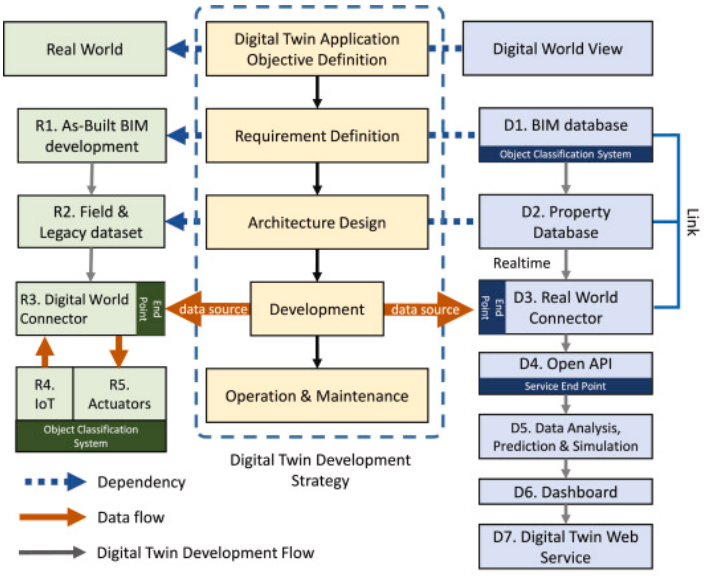

Fig. 2.3: Digital twin unmanned storage real-time mapping process logic diagram.

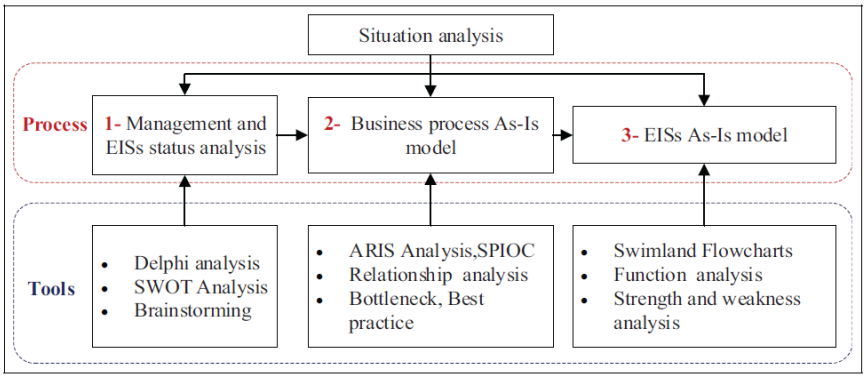

Fig. 2.4: Flow of resource efficiency optimization analysis.

2.3.1. Data analysis and prediction. The method of artificial neural network is adopted. Its input is based on the inbound and outbound commodity data collected by the digital twin data center, including the number of orders, order lines, received quantity, shipment quantity, inventory, dismantling amount, SKU and equipment status, etc., select the data related to the number of hidden layers, and divide it into training data, verification data and test data, the ratio of the three is about 7:1.5:1.5. The AUC value is used to determine the training effect, usually in the range of 0.5-1. The closer the value is to 1, the better the prediction effect of random judgment is. Combined with the collected data, the unmanned warehouse based on multidimensional information is scheduled, and its potential energy efficiency problems are fed back.

2.3.2. Automatic facility resource configuration. The automated system optimization process includes cluster analysis for device resource efficiency, and the generated unmanned warehouse must be encapsulated before it can be modeled to interact with deep reinforcement learning based on the Python language [12]. The unmanned warehouse model then takes the required form data from the cloud, executes the form and feeds

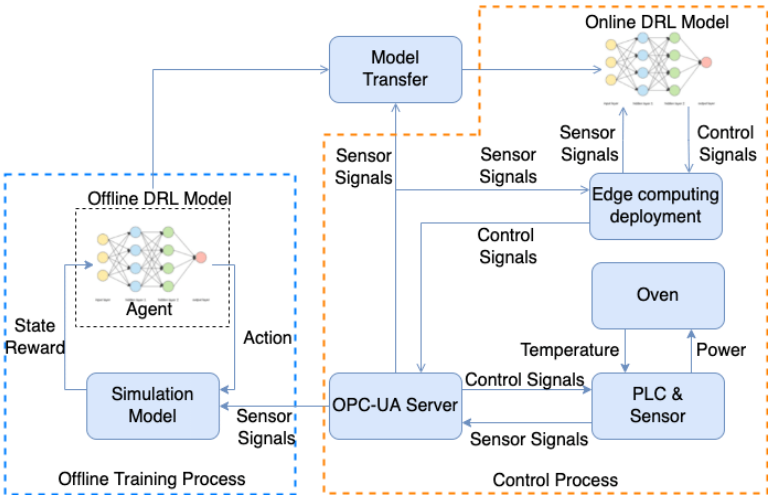

Fig. 2.5: Deep Reinforcement Learning Deployment Framework.

it back to the current state function. After obtaining the state matrix, the model makes decisions and operations. When the algorithm reaches the next determined time point, the algorithm will feed the current income and the state information of the next time point into the deep enhancement model. Finally, a trained deep reinforcement learning model for optimizing unmanned storage resources is obtained. The specific operation process is shown in Figure 2.5 (the picture is quoted in Using Deep Reinforcement Learning for Zero Defect Smart Forging).

2.3.3. Feedback of optimization results. Under the HTML architecture, the proven deep reinforcement learning mode is configured on Linux as an HTTP server. The jar bundle is configured in the cloud computing to access it as an API. In the implementation process, the data is collected, processed and integrated into the required data and uploaded to the cloud database [13]. In this paper, behavioral decisions are captured and fed back based on the deep reinforcement learning method of cloud computing and API models of the unmanned warehouse. Finally, the verified data is sent and returned to the terminal of the service layer. The user can see the optimal model and related parameters through the intuitive display interface. This makes the resource allocation of unmanned warehouses more scientific and reasonable.

3. Adaptive resource optimization method based on blockchain segmentation. The PPO method is a solid deep incentive learning method introduced by OpenAI in 2017, which is superior to other robust deep incentive learning methods in terms of sampling complexity [14]. By setting the trusted range, the method has an adaptive solid ability to avoid errors. Some scholars proposed adopting the PPO algorithm to adapt to the mapping error of digital twin model of the sheet metal assembly line to obtain the best clamping position. Currently, the edge data processing method based on the PPO method has a severe mapping error between the boundary twin and the physical network, and there are no corresponding research results. This paper studies the two-layer PPO algorithm to process different data types (Figure 3.1).

This paper presents a multi-layer PPO algorithm based on multiagent PPO. Block Administrator A uses a single PPO policy. Each K base station and block manager observed the existing dual-layer digital twin and imported the observations into the PPO neural network [15]. Finally, the boundary twin model was used to verify the output. Finally, the verified algorithm is optimized to the corresponding physical node. Compared with the conventional PPO method, this project proposes a dual multi-layer PPO method so that K BS and block administrators can obtain the data they need simultaneously, thus reducing the resource consumption required by manual intervention.

3.1. Application of two-layer PPO algorithm in digital twin. A two-level Markov decision model is constructed, and the model's state space, behavior space and reward function are studied.

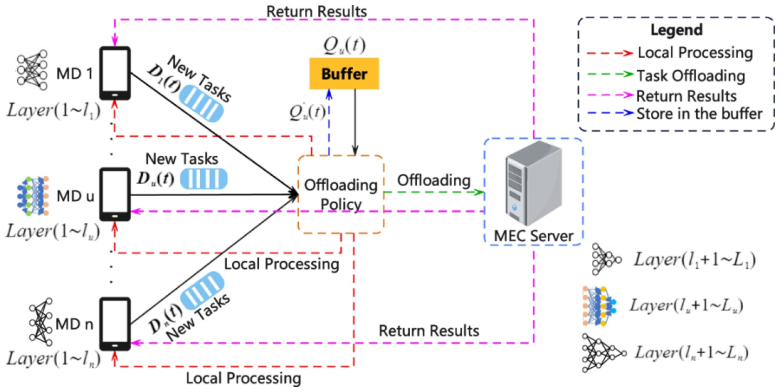

Fig. 3.1: PPO algorithm based on Downlink.

3.1.1. Phase space. At decision time $t(t = 1, 2, \ldots)$ there is a κ BS for maintaining the AP twin mode state of the local data sharing link. The algorithm includes the signal-to-noise ratio Θ_{n,j_k} of intelligent terminal n to $APjK$ and the signal-to-noise ratio $\Theta_{j\kappa,j\kappa'}$ of $APj\kappa$ to $APj\kappa'$, and $j_\kappa, j_\kappa{}' \in \mathfrak{J}_\kappa, j_\kappa \neq j_\kappa{}', n \in N_\kappa^\alpha, \alpha \in \alpha$. The state space of the κ base station is shown below

$$R_{\kappa,t}^z = \left[\Theta_{n,j_\kappa}, \Theta_{j_\kappa,j_{\kappa'}}\right]$$

K BS, whose state space is as follows

$$R_t^z = \left\{R_{1,t}^z, \cdots, R_{\kappa,t}^z, \cdots, R_{K,t}^z\right\}$$

In block administrator a, save the signal interference noise ratio $\Theta_{b,a}, \Theta_{a,\beta}, \Theta_{\beta,\beta'}, \Theta_{\beta,a}$ of the multicast transmission subchannel of the authentication node twin mode of the block, the maximum available computing resources $g_h^{\max}$ of the block manager and the authentication node, and $a, \beta, \beta' \in \beta, a \neq \beta, a \neq \beta', \beta \neq \beta'$. The block size is $R_{A,b}$, the local access is ASb and the block manager is a. Then, the state space of the block manager a can b expressed as:

$$\mid R_t^\beta = [R_{A,b}, \Theta_{b,a}, \Theta_{a,\beta}, \Theta_{\beta,\beta'}, g_h^{\max}]$$

3.1.2. Action space. The decision parameters of each node must be modified appropriately to meet the characteristics of time variability and maximize the benefit of K base stations and administrators of each layer [16]. The connection vector η between nodes and the bandwidth resource configuration vector Q_z of A-nodes are regulated in the local data sharing link. The local base station access vector is λ. The bandwidth resource configuration vector of the block manager and the parity node is Q_β. The resource allocation vector of the block manager and the parity node is g_β. In this way, the behavior space for optimal configuration of κ BS at the decision time t can be expressed as

$$H_{\kappa,t}^z = [\eta, Q_z]$$

Thus, the optimal behavior space for the utility of K BS can be expressed as

$$H_t^z = \left\{H_{1,t}^z, \cdots, H_{\kappa,t}^z, \cdots, H_{K,t}^z\right\}$$

In addition, the operation space for optimizing the utility of the regional manager a can be expressed as

$$H_t^\beta = [\lambda, Q_\beta, g_\beta]$$

3.1.3. Reward function. The constraints of C1 – C7 must be verified in the operation of Layer 2PPO, so the following real-time reward function r_t is proposed in this paper. Here's $r_t^z = \sum_{\kappa \in K} K_{BS_x} \gamma_t^\beta = K_a^\beta$. If C1-C7 constraints cannot be met at the same time, it means that the current optimal strategy is not effective [17]. To prevent invalid decisions, the immediate return is set to 0 .

3.2. Dual-layer PPO algorithm principle. Combining deep neural networks with reinforcement learning solves the constructed two-layer Markov decision problem. The two-layer PPO method obtains the best-determining variable ξ^* by establishing the best parameters of the artificial neural network. This maximizes the average rate of return in formula

$$S(\xi) = \mathrm{E}_{\delta\sim\pi_\xi(\cdot r)}\left[\sum_{t=1}^{T} 1 r_t\right]$$

$0 \leq 1 \leq 1$ stands for discount factor. E stands for random sampling based on transformation order δ. The expected value of the immediate return is found given the strategy π_ξ, and the state $r.\delta$ represents the sequence of conditions and behavior changes at the corresponding time point t, which is $\delta = \{R_1^z, R_1^\beta, H_1^z, H_1^\beta, \cdots, R_t^z, R_t^\beta, H_t^z, H_t^\beta\}$. PPO is a reinforcement learning method that uses new strategy gradients and confidence intervals. The network of actors accepts the current situation of the actors and makes decisions accordingly [18]. The confidence interval method is used to dynamically adjust the parameters in the network so that the network has adaptive solid ability and good convergence. The loss function of the update process of the Actor-network parameter ξ_k^a of the κBS is expressed as

$$J\left(\xi_\kappa^a\right) = \min\left(\sigma_t\left(\xi_k^a\right) H_t, \mathrm{clip}\left(\sigma_t\left(\xi_\kappa^a\right), 1-\pi, 1+\pi\right) H_t\right)$$

$\sigma_t\left(\xi_k^a\right)$ indicates the updating range of network parameters. H_t represents the dominance function, which reflects the decision generated by the current network parameters. Compared to other possible decisions, $H_{k,t}^z$ is of superior value. $\pi \in (0,1)$ is the parameter that determines the upper and lower boundary $(1-\pi, 1+\pi)$ of the PPO algorithm's confidence range. clip$(\cdot)$ function is used to constrain $\sigma_t\left(\xi_k^a\right)$, so it has adaptive solid ability and convergence. The definition of $\sigma_t\left(\xi_k^a\right)$ is

$$\sigma_t\left(\xi_\kappa^a\right) := \frac{\pi_{\xi_k^a}\left(H_{\kappa,t}^z \mid R_{\kappa,t}^z\right)}{\pi_{\xi_{k,ald}^a}\left(H_{\kappa,t}^z \mid R_{\kappa,t}^z\right)}$$

ξ_κ^a represents the network parameters that have been updated. ξ_κ^a,old is the network parameter before the upgrade. H_t is represented in formula (3.10). If the resulting decision $H_{\kappa,t}^z$ gets a better-expected return, then $H_t > 0$ is the opposite of $H_t < 0$. The dominant function H_t is defined in this way

$$H_t = \delta_t^z + \omega\delta_{t+1}^z + \cdots + \omega^{T-t+1}\delta_{T-1}^z$$

$\omega \in [0,1]$ stands for discount factor. δ_t^z represents the time error of a single step as defined below

$$\delta_t^z = r_{t+1}^z + \omega\rho\left(R_{t+1}^z\right) - \rho\left(R_t^z\right)$$

$\rho(\cdot)$ represents the Critic network's estimated reward for deciding $H_{\kappa,t}^z$. $r_{t+1}^z + \omega\rho\left(R_{t+1}^z\right)$ represents the sum of the immediate return r_{t+1}^z and the expected return of the Critic network corresponding to decision $H_{\kappa,t}^z$. The Critic neural network takes the change of the mean value of δ_t^z as its loss function. The weight ξ_k^z in the model is modified so that the cost function is maximum and the reward $\rho(\cdot)$ obtained by the algorithm is more accurate. The two different control strategies adopt the method of optimal learning rate to obtain the best network parameters $\xi_k^{a^*}$ and $\xi_k^{z^*}$.

4. System inspection. The project takes Z Company as an example to develop a digital twin unmanned warehouse system. According to the number of purchases made by the company between the first quarter of 2022 and the second quarter of 2023, they are classified and counted every week. The artificial neural network is used to analyze the actual production situation, and the AUC of 0.9245 is obtained, proving the method's effectiveness. This project adopts an A2C algorithm based on deep reinforcement learning for optimization [19]. The learning rate parameter was set to 1×10^{-6}, the simulation time step was set to 5min, the number of steps for each model training was 1000, and the total training step length was 5×10^6.

The tests compared to the inventory data are shown in Figure 4.1. Finally, the method is compared with those often used in unmanned warehouses. Higher rewards can be obtained through the optimal allocation

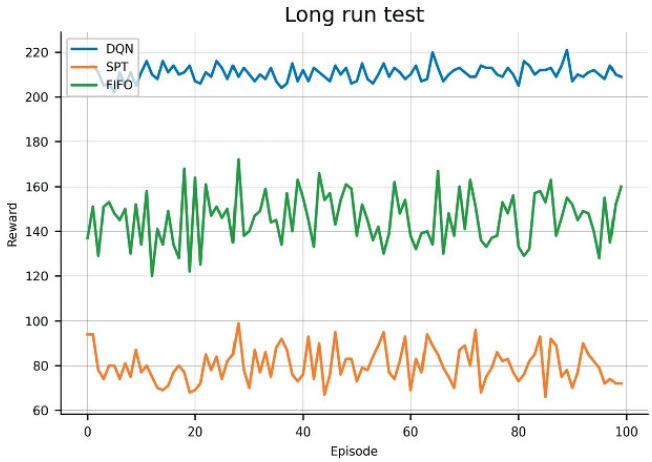

Fig. 4.1: Comparison of reward value, resource allocation and process time before and after optimization.

strategy. The deep enhancement method is adopted to optimize the design of forklift trucks in AGV, purchase area and loading and unloading area. The deep reinforcement learning method is used to configure the system resources dynamically, and the system's running speed is shortened from 26 minutes to 24.5 minutes. The time required to ship has been reduced from 3.6 points to 3.32 points. The material retention time in the warehouse was reduced from 44.21 minutes to 41.24 minutes. Through the dynamic resource adjustment of the system, the utilization rate and running speed are improved.

The best data information is returned to the data service system and is constantly adjusted to improve the model in the future. At the same time, these data will also be fed back to the data-sharing platform on the service side, and the decision maker can scientifically and reasonably allocate the corresponding resources according to the model and parameters on the visual interface.

5. Conclusion. A resource optimization method for unmanned warehouse systems based on deep reinforcement learning is proposed. This project uses simulation software to build a training environment for deep reinforcement learning. The model has effectively interacted with the production platform to realize the effective management of the authentic warehouse. An interactive model of an unmanned warehouse with man-machine interface is established. This project realizes collaborative optimization of unmanned warehouses based on cloud computing. This method has achieved good results in the actual operation of Z Company. The results prove the practicability of the proposed model, algorithm and prototype system.

Acknowledgement. The project is supported by the Scientific Research Project of Liaoning University of International Business and Economics(2023XJLXZD01).

REFERENCES

[1] Park, K. T., Jeon, S. W., & Noh, S. D. (2022). Digital twin application with horizontal coordination for reinforcement-learning-based production control in a re-entrant job shop. International Journal of Production Research, 60(7), 2151-2167.
[2] Bao, Q., Zheng, P., & Dai, S. (2024). A digital twin-driven dynamic path planning approach for multiple automatic guided vehicles based on deep reinforcement learning. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 238(4), 488-499.
[3] Yang, W., Xiang, W., Yang, Y., & Cheng, P. (2022). Optimizing federated learning with deep reinforcement learning for digital twin empowered industrial IoT. IEEE Transactions on Industrial Informatics, 19(2), 1884-1893.
[4] Shen, G., Lei, L., Li, Z., Cai, S., Zhang, L., Cao, P., & Liu, X. (2021). Deep reinforcement learning for flocking motion of multi-UAV systems: Learn from a digital twin. IEEE Internet of Things Journal, 9(13), 11141-11153.
[5] Badakhshan, E., & Ball, P. (2023). Applying digital twins for inventory and cash management in supply chains under physical and financial disruptions. International Journal of Production Research, 61(15), 5094-5116.
[6] Zhang, K., Cao, J., & Zhang, Y. (2021). Adaptive digital twin and multiagent deep reinforcement learning for vehicular edge computing and networks. IEEE Transactions on Industrial Informatics, 18(2), 1405-1413.

[7] Park, K. T., Son, Y. H., & Noh, S. D. (2021). The architectural framework of a cyber physical logistics system for digital-twin-based supply chain control. International Journal of Production Research, 59(19), 5721-5742.
[8] Goodwin, T., Xu, J., Celik, N., & Chen, C. H. (2024). Real-time digital twin-based optimization with predictive simulation learning. Journal of Simulation, 18(1), 47-64.
[9] Alexopoulos, K., Nikolakis, N., & Chryssolouris, G. (2020). Digital twin-driven supervised machine learning for the development of artificial intelligence applications in manufacturing. International Journal of Computer Integrated Manufacturing, 33(5), 429-439.
[10] Lee, J., Azamfar, M., Singh, J., & Siahpour, S. (2020). Integration of digital twin and deep learning in cyber-physical systems: towards smart manufacturing. IET Collaborative Intelligent Manufacturing, 2(1), 34-36.
[11] Marmolejo-Saucedo, J. A. (2022). Digital twin framework for large-scale optimization problems in supply chains: a case of packing problem. Mobile Networks and Applications, 27(5), 2198-2214.
[12] Ren, Z., Wan, J., & Deng, P. (2022). Machine-learning-driven digital twin for lifecycle management of complex equipment. IEEE Transactions on Emerging Topics in Computing, 10(1), 9-22.
[13] Rolf, B., Jackson, I., Müller, M., Lang, S., Reggelin, T., & Ivanov, D. (2023). A review on reinforcement learning algorithms and applications in supply chain management. International Journal of Production Research, 61(20), 7151-7179.
[14] Sun, W., Xu, N., Wang, L., Zhang, H., & Zhang, Y. (2020). Dynamic digital twin and federated learning with incentives for air-ground networks. IEEE Transactions on Network Science and Engineering, 9(1), 321-333.
[15] Wang, J., Li, X., Wang, P., & Liu, Q. (2024). Bibliometric analysis of digital twin literature: A review of influencing factors and conceptual structure. Technology Analysis & Strategic Management, 36(1), 166-180.
[16] Xu, H., Wu, J., Li, J., & Lin, X. (2021). Deep-reinforcement-learning-based cybertwin architecture for 6G IIoT: An integrated design of control, communication, and computing. IEEE Internet of Things Journal, 8(22), 16337-16348.
[17] Dai, Y., Zhang, K., Maharjan, S., & Zhang, Y. (2020). Deep reinforcement learning for stochastic computation offloading in digital twin networks. IEEE Transactions on Industrial Informatics, 17(7), 4968-4977.
[18] Sun, W., Lei, S., Wang, L., Liu, Z., & Zhang, Y. (2020). Adaptive federated learning and digital twin for industrial Internet of things. IEEE Transactions on Industrial Informatics, 17(8), 5605-5614.
[19] Ivanov, D., & Dolgui, A. (2021). A digital supply chain twin for managing the disruption risks and resilience in the era of Industry 4.0. Production Planning & Control, 32(9), 775-788.

Edited by: Hailong Li
Special issue on: Deep Learning in Healthcare
Received: Apr 19, 2024
Accepted: May 27, 2024