Research on Omni-Channel Inventory Management Based on Deep Reinforcement Learning

Changhui Ma*, Jiaqi Ding

Information Management and Information System, Lanzhou University, Lanzhou, China, 730000

*Corresponding author: 320220915840@lzu.edu.cn

Abstract. This study aims to address the complexity of inventory management in an omni-channel retail environment by proposing an intelligent inventory management model based on Deep Reinforcement Learning (DRL). The research objective is to develop a comprehensive framework that integrates various stages of the retail supply chain to enhance operational efficiency and customer service. By introducing Deep Q Networks (DQN) and a multi-agent collaboration framework, the model achieves the coordinated optimization of various stages, including inventory allocation, demand forecasting, order fulfillment, and return processing. The research innovatively integrates the return management mechanism with inventory replenishment strategies, proposing dynamic inventory allocation and real-time decision-making mechanisms that significantly enhance the efficiency and flexibility of omni-channel inventory management. Experimental results show that this model can significantly improve service levels (from 85% to 96.8%) in dynamic environments, optimize inventory costs, and enhance the operational capabilities of enterprises under complex market conditions. The study not only enriches the theoretical framework for omni-channel retail inventory management but also provides enterprises with a practical intelligent decision-making tool, which is of great significance for promoting the digital transformation of retail businesses.

Keywords: Omni-channel Retail, Inventory Management, Deep Reinforcement Learning, Multi-agent Systems, Dynamic Optimization.

1. Introduction

With the development of e-commerce and changes in consumer behavior, over 80% of retailers have transitioned to omnichannel retail models [1]. In the omnichannel environment, consumers can access products through multiple channels, while retailers face complex challenges such as channel coordination, inventory allocation, demand forecasting, and return management[2]. Traditional inventory management methods are limited in high-dimensional, dynamic environments. Deep Reinforcement Learning (DRL), with its ability to handle complex decision-making problems, has become a powerful tool for addressing inventory management in omnichannel retailing. This paper reviews the research progress of DRL applications in inventory management for omnichannel retailers.

Omnichannel inventory management faces multiple challenges. Goedhart et al. (2022) studied inventory allocation problems using the Markov Decision Process (MDP) and demonstrated that dynamic allocation strategies outperform static strategies, providing theoretical support for the use of dynamic decision-making methods like DRL[3]. Goedhart et al. (2023) further studied the impact of returns, finding that an increase in online return rates reduces retailer profits, emphasizing the importance of accurately modeling return flows[4].

Vanvuchelen et al. (2024) developed a horizontal allocation strategy using Proximal Policy Optimization (PPO), improving service level fairness in resource-constrained environments[5]. Yavuz and Kaya (2024) found that as problem complexity increases, traditional dynamic programming faces the "curse of dimensionality," while Deep Q-Learning (DQL) achieved more than 95% approximation accuracy and significantly reduced solution times[6].

Liang et al. (2024) proposed the DCEA-DQN algorithm, combining deep Q networks with evolutionary algorithms, using a double DQN architecture to handle dynamic constraint problems and achieving optimal performance in 80% of the test scenarios[7].

Mohamadi et al. (2024) used the Advantage Actor-Critic (A2C) algorithm to optimize perishable goods inventory allocation, improving the service level from 58% to 86% while almost eliminating waste, demonstrating the effectiveness of DRL in continuous state and action spaces[8].

Kaynov et al. (2024) proposed a multi-discrete action distribution method for the single warehouse, multiple retailer (OWMR) system, reducing the neural network output layer from exponential (A^K) to linear ($A \times K$)and developed a random order allocation strategy to improve learning efficiency[9].

The Internet of Things (IoT) and computer vision technologies provide high-quality real-time data for DRL. Villegas-Ch et al. (2024) developed a computer vision-based inventory platform that improved inventory counting accuracy to 91%, providing reliable data support for DRL decision-making[10]. Furthermore, Liu et al. (2024) proposed digital twin technology, which simulates inventory flows and customer purchasing behaviors across different channels, providing rich virtual environments for DRL model training[11].

In response to the above-mentioned limitations, this study proposes an optimization decision model for omnichannel retail inventory management based on DRL technology, with the following innovative approaches:

Comprehensive Integrated Omnichannel Inventory Management Framework: This research overcomes the limitation of focusing solely on single or few inventory management issues, proposing a comprehensive integrated omnichannel inventory management framework. This framework organically combines inventory allocation, demand forecasting, order fulfillment, and return management, forming a collaboratively optimized system. By establishing a unified state space and designing reward functions, this model achieves coordinated decision-making across functions, solving the problem of partial optimization in previous studies such as those by Goedhart et al. (2022) and Vanvuchelen etal. (2024).

This study uses a multi-agent reinforcement learning approach to solve the order fulfillment decision optimization issues found in Yuan and Wang (2025). The system includes warehouse agents, store agents, and scheduling agents, dynamically optimizing order fulfillment paths based on real-time inventory status, order characteristics, and distribution costs. The system adopts a hierarchical reinforcement learning architecture, with high-level agents handling global resource allocation and low-level agents executing specific order processing, overcoming the limitations of static rules and simplified models in previous research. This approach is particularly suitable for complex network environments involving multiple warehouses and retail points.

Through the above innovative methods, this research provides retailers with a feasible solution for omnichannel inventory management and offers new theoretical frameworks and practical guidance for future research. Experimental results show that compared to traditional methods and existing DRL approaches, the proposed model achieves significant improvements in key indicators such as total cost, service levels, and resource utilization, demonstrating its effectiveness and superiority in omnichannel retail inventory management.

2. Model Design and Experimental Methods

2.1. Simulation System Architecture

This study constructs a simulation system for inventory management that mimics an omnichannel retail environment. By combining deep reinforcement learning with a multi-agent system, the system achieves intelligent optimization of inventory decisions. The system architecture is shown in Fig. 1.

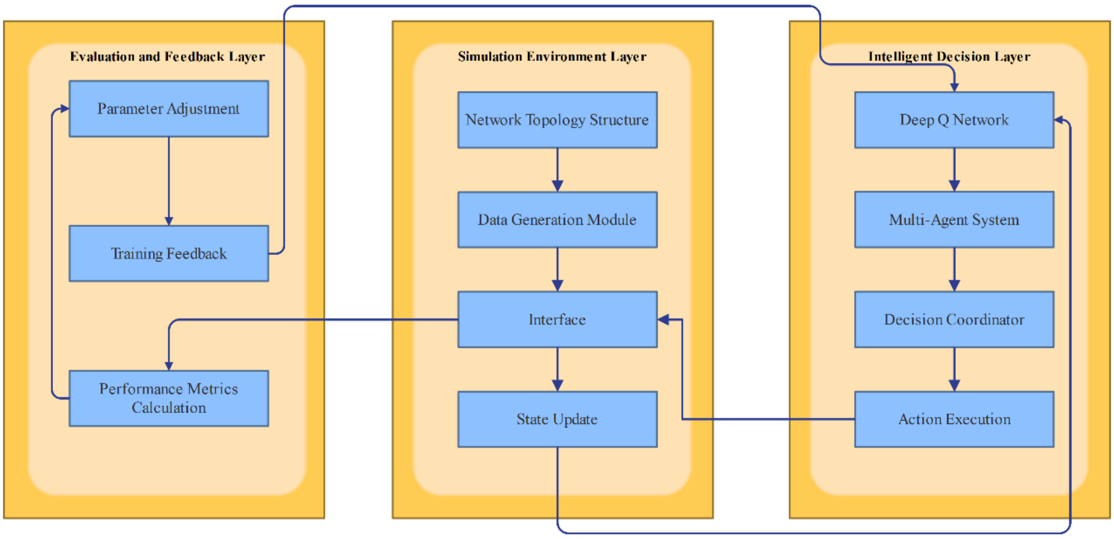

Fig. 1. System Architecture

2.1.1 System Overall Design

The system adopts a three-tier architecture, with the functionalities of each layer as follows:

(1) Simulation Environment Layer

Network Topology: Simulates a distribution network consisting of one central distribution center (DC) and ten retail stores.

Interaction Interface: Provides standardized interfaces for state observation and action execution.

State Updates: Real-time updates of inventory levels, order status, and cost data.Environment Parameters: Includes basic parameters such as storage capacity, transportation time, and processing costs.

(2) Intelligent Decision Layer

Deep Q Network: Handles high-dimensional state spaces and outputs decision actions.Multi-Agent System: Includes three types of agents: ordering, shipping, and return processing agents.Decision Coordinator: Coordinates the decision outcomes of multiple agents.Action Execution: Transforms decisions into specific operational instructions.

(3) Evaluation and Feedback Layer

Performance Metrics: Real-time calculation of key metrics such as service levels and inventory costs.Parameter Adjustment: Dynamically adjusts training parameters based on performance evaluation results.Training Feedback: Provides guidance information for decision-making optimization.

2.1.2 Data Generation Mechanism

The simulation system generates various types of data required for experiments through the following mechanisms:

(1) Demand Data Simulation

$$\text{Demand}(t) = \text{Base Demand} + \text{Seasonal Fluctuations} + \text{Random Disturbances} \tag{1}$$

Where:Base Demand: The average level set based on historical sales data.

Seasonal Fluctuations: Simulated periodic changes using a sine function, with an amplitude of 20% of the base demand.

Random Disturbances: Normally distributed with a mean of 0 and a standard deviation of 10% of the base demand.

(2) Return Behavior Generation

$$\text{Returns}(t) = \lambda \times \text{Demand}(t) \tag{2}$$

Where: λ is the return rate, set between 5% and 15% based on product categories.The return rate for different stores has a random fluctuation of ±2%.

(3) Cost Data Construction

(a)Inventory Holding Cost: Daily warehousing fees per unit of product.

$$\text{Inventory Cost} = \text{Inventory Units} \times \text{Unit Holding Cost} \times \text{Holding Days} \tag{3}$$

(b)Transportation Cost: Tiered pricing based on distance and quantity.

(c)Return Processing Cost: Includes logistics and restocking costs.

$$\text{Return Cost} = \text{Base Processing Fee} + \text{Unit Processing Cost} \times \text{Return Quantity} \tag{4}$$

Through this layered architecture design and data generation mechanism, the simulation system can effectively simulate inventory management scenarios in an omnichannel retail environment, providing a reliable experimental platform for subsequent algorithm validation. The system's modular design also facilitates parameter adjustments and functional expansions, enhancing the flexibility and controllability of the experiments.

2.2. Intelligent Optimization Model

This study constructs an intelligent optimization model by combining deep reinforcement learning and a multi-agent system. Through the collaborative decision-making of multiple specialized agents and the learning capabilities of the Deep Q Network (DQN), intelligent optimization of omnichannel inventory management is achieved.

2.2.1 Multi-Agent Collaborative Framework

The system designs three types of specialized agents, each responsible for ordering, shipping,and return processing tasks. Each agent learns the optimal strategy through a shared experience pool and collaborates on decision-making via a decision coordinator. The collaboration between agents is achieved through the following mechanisms:

(1) Experience sharing updates, represented by a mathematical formula:

$$Q(s,a) \leftarrow Q(s,a) + \alpha\left[r + \gamma \max_{a'} Q(s',a') - Q(s,a)\right] \tag{5}$$

(2) Decision coordination, where the final decision (a_final) is calculated as a weighted average of decisions from N agents:

$$a_{\text{final}} = \frac{1}{N}\sum a_i \tag{6}$$

2.2.2 Deep Q Network Structure

An improved Deep Q Network is used to handle decision-making in high-dimensional state spaces. The network structure consists of:

(1) State Input Layer: Processes environmental information such as inventory levels and demand.

(2) Hidden Layer: Extracts features and computes Q-values.

(3) Action Output Layer: This section explains how specific decision actions are generated. It describes stability enhancements through experience replay and target network mechanisms:

(a) Experience replay sampling: Defined as "experience = (state, action, reward, next_state)" with random sampling from a replay buffer.

(b) Target Q-value computation:

$$Q_{\text{target}} = \text{reward} + \gamma \cdot \max Q'(\text{next\_state}, \text{a}') \tag{7}$$

2.2.3 Reward Mechanism Design

This section details a multi-objective reward function designed to balance service levels and cost control:

Combined reward calculation:

$$R = \alpha \cdot \text{ServiceLevel} - \beta \cdot \text{InventoryCost} - \gamma \cdot \text{ReturnCost} \quad (8)$$

(1) ServiceLevel: Order fulfillment rate
(2) InventoryCost: Inventory holding cost
(3) ReturnCost: Return processing cost
(4) α, β, γ : Weight coefficients that balance these different objectives

Through this multi-layered intelligent optimization model design, the system is capable of:(1)Achieving collaborative decision-making for multiple tasks,(2)Handling complex state spaces,(3)Balancing multiple optimization objectives,(4)Maintaining training stability

3. Experimental Results Analysis

This section evaluates the performance of the multi-agent reinforcement learning-based inventory management model in an omnichannel retail scenario. By analyzing the training process and final outcomes, the effectiveness of the method is validated.

3.1. Experimental Results Analysis

3.1.1 Training Process Evaluation

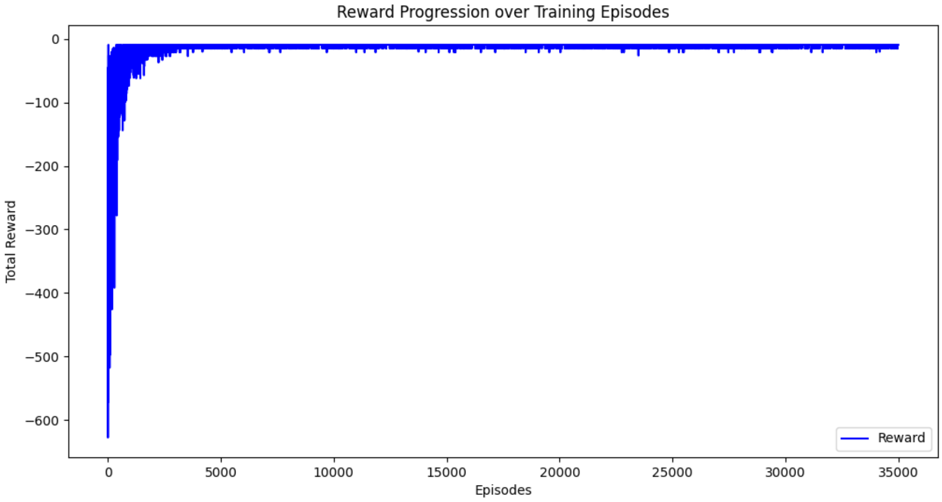

Fig. 2. Reward Progression

Fig. 3. Service Level

In terms of reward progress(Fig. 2), the model shows a clear learning process. In the initial phase (0-200 episodes), significant learning effects were observed, with the reward value increasing rapidly.

In the mid-phase (200-600 episodes), the model entered a stable optimization stage, and fluctuations gradually decreased. In the final phase (600-1000 episodes), the reward value stabilized at a high level, indicating that the strategy was approaching optimality. Meanwhile, the service level steadily increased from an initial 85% to 96.8%, reflecting the optimization effects in order fulfillment(Fig. 3).

3.1.2 Inventory Optimization Effects

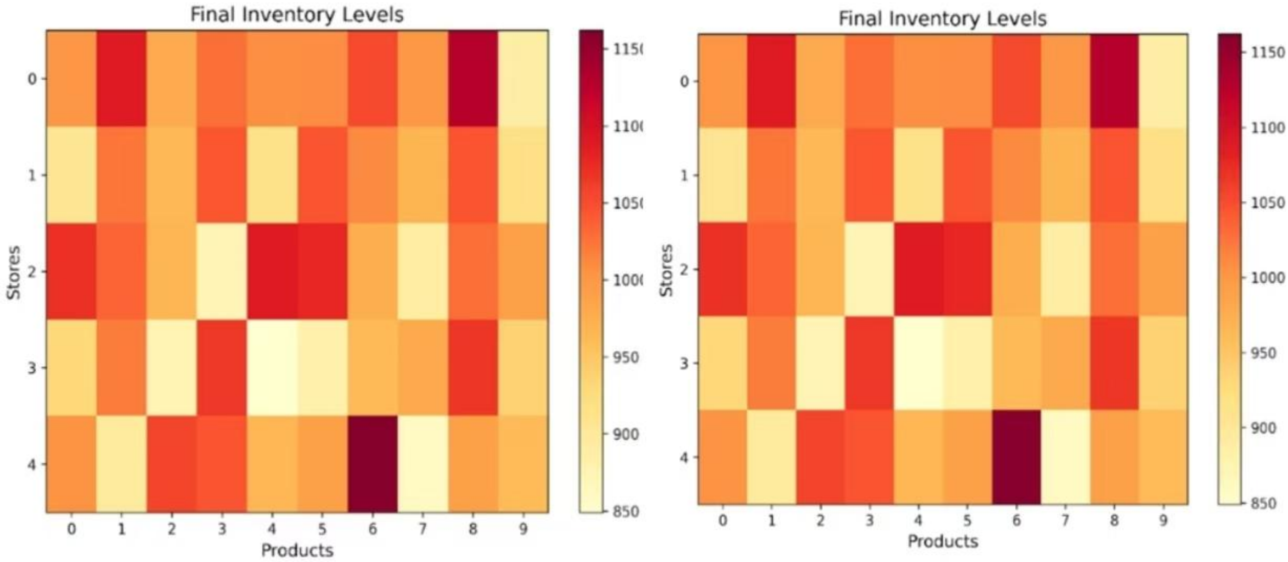

Fig. 4. Heatmap of Comparison of Inventory Optimization Before and After

According to Fig. 4,the improvement in inventory distribution was significant. Compared to the initial state where the inventory distribution was uneven, with excess stock and stockouts, the optimized inventory levels became more balanced, and scheduling became more reasonable. Inventory turnover improved significantly. The model achieved performance improvements across multiple dimensions: inventory costs were reduced by minimizing unnecessary stockpiling, return processing efficiency was enhanced through optimized reuse strategies, and order response speed was improved by optimizing delivery path selection.

3.2. Key Findings

In terms of multi-agent collaboration, the experimental results show that the agents achieved effective collaboration through information sharing, and the decision coordination mechanism ensured global optimality. The system as a whole exhibited good stability and scalability. In terms of inventory management efficiency, the model demonstrated strong dynamic adjustment capabilities, adapting to demand fluctuations. It also showed high accuracy in predictions, effectively reducing stockout risks. The model excelled in cost control, successfully balancing service levels and operational costs.

4. Conclusions

This research proposes an intelligent inventory management model for omni-channel retail based on Deep Reinforcement Learning (DRL). By integrating a multi-agent collaborative framework with improved DRL technology, the model successfully addresses the challenges of inventory allocation, demand forecasting, order fulfillment, and return processing in a unified approach. Experimental results demonstrate significant improvements in key performance indicators, with service levels increasing from 85% to 96.8% while simultaneously optimizing inventory costs and distribution.

The main contributions include: (1) a comprehensive integrated framework that achieves coordinated optimization across multiple functions; (2) an innovative return-inventory joint optimization model; (3) an intelligent order fulfillment decision system based on multi-agent reinforcement learning; (4) a novel channel coordination mechanism featuring shared inventory pools; and (5) technological improvements to DRL implementation that enhance real-world applicability.

Future research should focus on incorporating more sophisticated demand forecasting models, integrating additional data sources such as IoT devices, and validating the model in real-world retail environments. This study not only enriches the theoretical framework for omni-channel retail inventory management but also provides retailers with a practical intelligent decision-making tool to improve competitiveness in increasingly dynamic marketplaces.

References

[1] Yuan, D., & Wang, Y. Sustainable supply chain management: A green computing approach using deep Q-networks[J]. Sustainable Computing: Informatics and Systems, 2025, 45: 101063.

[2] Srinivas, D., Twinkle, P. D. Sawant, F. Rabby, M. S. Vasu, & A. Roniboss. Supply Chain Optimization: Machine Learning Applications in Inventory Management for E-commerce[C]. Proceedings of the 7th International Conference on Contemporary Computing and Informatics (IC3I), Greater Noida, India, 2024: 457-461.

[3] Goedhart, J., Haijema, R., & Akkerman, R. Inventory rationing and replenishment for an omni-channel retailer[J]. Computers & Operations Research, 2022, 140: 105647.

[4] Goedhart, J., Haijema, R., & Akkerman, R. Modelling the influence of returns for an omni-channel retailer[J]. European Journal of Operational Research, 2023, 306(3): 1248-1263.

[5] Vanvuchelen, N., De Boeck, K., & Boute, R. N. Cluster-based lateral transshipments for the Zambian health supply chain[J]. European Journal of Operational Research, 2024, 313(1): 373-386.

[6] Yavuz, T., & Kaya, O. Deep reinforcement learning algorithms for dynamic pricing and inventory management of perishable products[J]. Applied Soft Computing, 2024, 163: 111864.

[7] Liang, Z., Yang, R., Wang, J., Liu, L., Ma, X., & Zhu, Z. Dynamic constrained evolutionary optimization based on deep Q-network[J]. Expert Systems with Applications, 2024, 249(Part B): 123592.

[8] Mohamadi, N., Akhavan Niaki, S. T., Taher, M., & Shavandi, A. An application of deep reinforcement learning and vendor-managed inventory in perishable supply chain management[J]. Engineering Applications of Artificial Intelligence, 2024, 127(Part B): 107403.

[9] Kaynov, I., van Knippenberg, M., Menkovski, V., van Breemen, A., & van Jaarsveld, W. Deep Reinforcement Learning for One-Warehouse Multi-Retailer inventory management[J]. International Journal of Production Economics, 2024, 267: 109088.

[10] Villegas-Ch, W., Maldonado Navarro, A., & Sanchez-Viteri, S. Optimization of inventory management through computer vision and machine learning technologies[J]. Intelligent Systems with Applications, 2024, 24: 200438.

[11] Liu, S., Wang, J., Wang, R., Zhang, Y., Song, Y., & Xing, L. Data-driven dynamic pricing and inventory management of an omni-channel retailer in an uncertain demand environment[J]. Expert Systems with Applications, 2024, 244: 122948.