Deep Reinforcement Learning with Convolutional Neural Networks for Optimizing Supply Chain Inventory Management

Sheng Wang, Fan Li*
College of Economics and Management, Jiaozuo University, Jiaozuo 454000, China
E-mail: li-fan188@outlook.com
*Corresponding author

Keywords: reinforcement learning, convolutional neural network, supply chain, inventory control, demand forecasting

Received: February 25, 2025

With the increasing complexity of global supply chains and frequent fluctuations in market demand, inventory management faces severe challenges, and traditional inventory control methods are difficult to meet the needs. This paper constructs an inventory control model that combines deep reinforcement learning (DRL) with convolutional neural networks (CNN). By defining the state space, action space and reward function, the Q-learning algorithm is used to optimize inventory decisions. At the same time, CNN is used to extract historical demand data features to improve the accuracy of demand forecasting. This study uses historical sales data from a medium-sized clothing retailer as a dataset, which contains sales, inventory, and replenishment records for the past 4 years. The model was trained for 500 episodes. This model was compared with the economic order quantity (EOQ) model, the periodic ordering model, and the simple moving average forecasting model as the benchmark model. The model's demand forecast error of 3.2% was measured on independent actual test data. The experimental results show that the model has a demand forecast error of only 3.2%, the total inventory cost is 14,500 yuan, the cost reduction rate is -22%, the average inventory turnover rate is 10.5 times, and the average out-of-stock rate is only 2.1%. All indicators are significantly better than the economic order quantity (EOQ) model and the periodic ordering model. The study proves that the model can effectively cope with demand fluctuations and uncertainties, optimize inventory management, and provide a new and effective method for supply chain inventory control.

Povzetek: Prispevek predstavlja model za upravljanje zalog, ki združuje globoko ojačitveno učenje in konvolucijske nevronske mreže ter dinamično optimizira napovedovanje povpraševanja in zalog v kompleksnih dobavnih verigah.

1 Introduction

In the complex environment of the global supply chain, inventory management has always been a critical and challenging issue. When managing inventory, modern enterprises often face the problem of how to ensure sufficient supply while avoiding excess inventory [1]. Excessive inventory not only brings high storage costs, but also may lead to capital occupation, thus affecting the company's liquidity; while insufficient inventory may cause out-of-stock, affect the continuity of the production line, and even lead to customer loss [2]. Accurately grasping this supply and demand balance is an arduous task in supply chain management. Especially driven by the wave of globalization and digitalization, market demand changes are becoming more frequent and complex, and the difficulty of enterprises in inventory control has increased accordingly [3].

For example, according to a report released by the International Federation of Robotics (IFR), the economic losses caused by global supply chain disruptions in 2021 exceeded US$4 trillion, of which improper inventory management was one of the main factors leading to the losses [4]. With the advancement of information technology and the increasing availability of data, traditional inventory control methods, such as economic order quantity (EOQ) and just-in-time (JIT), still have their role, but they are gradually becoming incapable of coping with the complex and dynamically changing market environment.

In current research, although reinforcement learning has applications in inventory management, the scalability of the model and its adaptability to extreme market fluctuations in dealing with complex supply chain scenarios with multiple products and suppliers are still insufficient. In addition, there is a lack of in-depth exploration in combining deep learning technology to improve the synergy between demand forecast accuracy and inventory control strategy optimization

This study aims to propose an innovative supply

chain inventory control framework based on reinforcement learning to solve some key problems in current supply chain inventory management" was modified to "This study aims to reduce the demand forecast error to less than 5%, reduce inventory costs by more than 20%, increase inventory turnover to more than 10 times, and control the out-of-stock rate to less than 3% by constructing an inventory control model that combines deep reinforcement learning with convolutional neural networks, so as to effectively deal with key issues such as demand fluctuations, supply uncertainties, and supply chain delays in complex supply chain environments.

Therefore, how to use smarter tools to optimize inventory management has become an important issue in academia and industry. In this context, reinforcement learning (RL), as a self-learning and optimization artificial intelligence technology, has begun to be widely used in supply chain management, especially in the optimization of inventory control strategies. Through the RL model, the supply chain can learn autonomously in a constantly changing environment, thereby making more accurate and flexible inventory decisions, providing a new way to solve problems that are difficult to deal with with traditional methods [5].

Reinforcement learning, as a branch of machine learning, has attracted widespread attention in supply chain management in recent years. Related research shows that RL can gradually find the best inventory management strategy by simulating the decision-making process [6]. For example, research shows that in the face of volatile demand, RL can effectively surpass the traditional EOQ model, achieve more accurate order quantity forecasts, and avoid inventory backlogs or out-of-stock phenomena [7].

In addition, inventory turnover is crucial for the efficient circulation of corporate funds and the improvement of operational efficiency. Traditional inventory management methods make it difficult to effectively improve inventory turnover while ensuring supply and avoiding inventory backlogs. Therefore, how to achieve a significant increase in inventory turnover in a complex supply chain environment has become one of the core concerns of this study.

However, despite the great potential of RL in inventory control, existing research still faces many challenges. First, the problem of balancing "exploration" and "utilization" in reinforcement learning often leads to slow convergence of the learning process or excessive sensitivity to the setting of hyperparameters in practical applications. In addition, many existing RL models are suitable for small-scale, single-link inventory optimization problems, but the scalability and application effect of RL in complex supply chains with multiple levels and suppliers are still unclear [8, 9].

How can deep reinforcement learning and convolutional neural networks be effectively combined to accurately predict demand in the supply chain inventory system? What is the optimal inventory control strategy under the influence of various uncertainties in the supply chain, such as demand fluctuations and supply delays? How to improve the overall efficiency of supply chain inventory management, including reducing inventory costs, increasing inventory turnover rate, and minimizing out - of - stock rate through the proposed model?

In addition, there is currently a lack of unified standards for the evaluation of RL in supply chain inventory control. Although many theoretical models have achieved good results in experimental environments, when they are applied in actual supply chains, they are often affected by external factors, such as supply chain uncertainty and market fluctuations, which makes the effects of these theoretical models not necessarily fully transformed into advantages in actual operations. Therefore, how to overcome these problems and make the practical application of RL in complex supply chains more efficient and feasible has become an important issue that needs to be solved in this field.

This study aims to propose an innovative supply chain inventory control framework based on reinforcement learning. By designing an RL model that can learn and optimize inventory decisions in a dynamic environment, this study aims to solve some key problems in current supply chain inventory management, such as demand fluctuations, supply uncertainty, and supply chain delays. Through the proposed RL model, this study will explore how to improve the accuracy of inventory management, reduce excess inventory and out-of-stock problems, and thus achieve more efficient inventory control in an uncertain environment.

Table 1: Key research on reinforcement learning in inventory management

Research	method	Performance Indicators	limitation
Study 1: [1]	[Use simple Q-learning algorithm and basic inventory rules to make inventory decisions]	[Inventory cost reduced by 15%, out-of-stock rate 8%]	[Only applicable to simple single-product supply chain scenarios, poor adaptability to complex

Research	method	Performance Indicators	limitation
			demand fluctuations]
Study 2: [2]	[Reinforcement learning model based on Deep Q Network (DQN), trained using historical sales data]	[Average inventory turnover increased by 8 times and inventory holding costs decreased by 12%]	[High data volume requirements, performance degradation when data is sparse, and long training time]
Study 3: [3]	[Using policy gradient algorithm to optimize inventory strategy, considering multi-stage supply chain]	[Service level reaches 90%, total operating cost decreases by 10%]	[Algorithm convergence is slow and hyperparameter adjustment is difficult]

Table 1 lists the specific names of past studies on reinforcement learning in inventory management. The methods describe in detail the specific algorithms and inventory management strategies used in each study. For example, Study 1 uses a simple Q-learning algorithm combined with basic inventory rules to make inventory decisions, so that readers can clearly understand the technical route of the study. The performance indicators clearly give the specific performance results achieved by each study through its method. For example, Study 1 achieved a 15% reduction in inventory costs and an 8% out-of-stock rate. These quantitative indicators help to intuitively compare the effectiveness of different studies. Limitations: point out the shortcomings of the methods or models used in each study. For example, Study 1 is only applicable to simple single-product supply chain scenarios and has poor adaptability when facing complex demand fluctuations. This provides a reference direction for the improvement and innovation of subsequent research, and also allows readers to have a more comprehensive understanding of the status of existing research. Through such a table summary, the similarities and differences of past related studies can be more clearly compared, highlighting the innovations and improvement directions of this study, making the literature review completer and more convincing.

The significance of this study is not only reflected in its theoretical contribution, but also has important practical value. Through the proposed RL optimization framework, enterprises can get rid of the limitations of traditional inventory control methods and adopt a more flexible and intelligent way to manage inventory. Unlike traditional static inventory models, RL models can automatically adjust inventory strategies according to real-time market changes, thereby improving the responsiveness and efficiency of the overall supply chain. This transformation will help reduce the operating costs of enterprises, improve inventory turnover, and enhance customer satisfaction.

From an academic perspective, this study will provide a new perspective for the application of reinforcement learning in supply chain management, especially for optimization applications in multi-level and complex supply chain environments. The results of this study will not only help fill the gap in the current academic community in this field, but will also lay the foundation for the wider application of RL in the supply chain in the future. In addition, the results of this study are of great reference significance to supply chain managers and policymakers in the industry, especially in terms of how to use advanced artificial intelligence technology to improve the efficiency and risk resistance of supply chain management.

This study innovatively combines deep reinforcement learning with convolutional neural networks to propose a new dynamic optimization framework for complex supply chain inventory management problems. Unlike previous methods that rely on a single technology or simple combination, this model uses convolutional neural networks to perform deep feature mining on historical demand data, providing more accurate state input for reinforcement learning and achieving intelligent and adaptive optimization of inventory decisions.

2 Literature review

2.1 Application of reinforcement learning in supply chain management

In the past few years, reinforcement learning (RL)

has gradually shown great potential in supply chain management, especially in the field of inventory control. Through autonomous learning and continuous adjustment of strategies, RL is able to optimize inventory decisions in dynamic and uncertain environments. The latest research in this field focuses on how to overcome the limitations of traditional methods and improve the efficiency and adaptability of inventory management.

For example, the application of Q-learning-based reinforcement learning models in multi-echelon supply chain systems has made significant progress. This type of model can repeatedly learn and find the optimal inventory control strategy by simulating factors such as orders, inventory, and demand in the supply chain. Unlike traditional EOQ (economic order quantity) and JIT (just-in-time) models, RL methods do not rely on pre-set rules, but instead continuously adjust inventory levels through interaction with the environment to achieve the goal of reducing costs and improving supply chain responsiveness. In addition, deep reinforcement learning (DRL), as an innovative technology that combines deep learning and reinforcement learning, has gradually demonstrated its advantages in inventory optimization problems in recent years. DRL can not only handle more complex supply chain environments, but also find more accurate inventory control strategies in changing demand and supply chain uncertainties [10, 11].

Although RL methods have strong adaptability and flexibility, current research still faces some challenges, especially in the application of complex supply chain environments with multiple suppliers and multiple products. Most existing RL models are based on a simplified single supply chain node and lack in-depth consideration of the multi-level structure and changing environment of the actual supply chain. This simplification limits the performance of RL models when dealing with large-scale supply chain networks. Therefore, how to design more complex and realistic RL models to better cope with inventory management problems in multi-level supply chain systems is still a major difficulty in current research [12, 13].

2.2 Optimization of reinforcement learning model and algorithm improvement

In the process of applying reinforcement learning to supply chain inventory management, model optimization and algorithm improvement have always been hot topics of research. On the one hand, how to accelerate the convergence speed of RL algorithms, and on the other hand, how to improve the stability and accuracy of models have become the core issues that researchers are concerned about. In order to meet this challenge, recent studies have proposed a variety of new reinforcement learning algorithms, such as the RL framework combined with policy gradient optimization methods and model predictive control (MPC) [14, 15].

Policy gradient-based algorithms usually rely on direct optimization of policies. Compared with traditional Q-learning, they can avoid the interference of value function estimation errors on the learning process. By gradually adjusting the policy, policy gradient-based RL algorithms can achieve better results in more complex inventory control problems [16]. In addition, with the improvement of computing power, algorithms such as deep Q network (DQN) have gradually become the mainstream method of RL in supply chain management [17]. These methods introduce neural networks to approximate the Q value function, enabling RL to handle high-dimensional state space and complex inventory management problems.

However, although these new algorithms have shown good performance in experiments, they still face many challenges in practical applications. For example, deep reinforcement learning often requires a large amount of training data and computing resources, while the data in actual supply chain environments is often incomplete or noisy, which limits the applicability of the algorithm in reality. Therefore, how to design more efficient and robust reinforcement learning algorithms and reduce dependence on data and computing resources remains a key issue that needs to be solved in this field [18].

2.3 Integration of reinforcement learning and other technologies

In modern supply chain management, RL does not operate in isolation. Many studies have begun to explore the combination of RL with other technologies to improve the efficiency and feasibility of inventory control. This integration is not limited to the combination with traditional algorithms, but also includes the combination with emerging technologies such as big data analysis, cloud computing, and the Internet of Things (IoT). This interdisciplinary integration has brought new ideas and methods to supply chain inventory management [19].

In recent years, research has begun to focus on the combination of RL and big data, providing more accurate demand forecasting and inventory management solutions through real-time monitoring and analysis of large-scale supply chain data. Through big data analysis, RL models can more accurately capture demand fluctuations and market changes in the supply chain, thereby formulating more personalized inventory strategies [20]. On the other hand, the rapid development of Internet of Things technology enables each link in the supply chain to obtain data in real time through sensors, which provides RL models with rich real-time information and further enhances the model's real-time decision-making ability [21].

In addition, the combination of RL and cloud computing has also shown great application potential. Through the cloud platform, each link in the supply chain

can obtain computing resources and storage space more flexibly, allowing the RL model to process larger-scale supply chain data and make more complex inventory optimization decisions. This cloud computing-based RL framework can achieve real-time scheduling and optimization of supply chain management, which is particularly suitable for global and highly dynamic supply chain environments [22].

Although these fusion methods provide more flexible and efficient inventory control solutions, their implementation still faces many challenges. Data privacy issues, technology integration issues, and high system complexity are all difficulties in current research. Therefore, how to solve these problems and achieve the best effect of technology fusion is an important direction for future research [23].

3 Research methods

3.1 Theoretical basis of the model

The inventory control model proposed in this paper combines deep reinforcement learning (DRL) with traditional inventory control strategies, and adopts a reinforcement learning framework based on Q-learning to dynamically optimize inventory decisions. Traditional inventory management methods often rely on static rules and preset models, and fail to effectively cope with the dynamic fluctuations and uncertainties of demand. To this end, this paper innovatively introduces a reinforcement learning algorithm to continuously adjust inventory decisions through model self-learning, so that the system can achieve dynamic optimization in a changing market environment [24].

We define the state space of the system S and action space A To describe the decision-making process of the inventory control problem. State space S It includes multiple dimensions, representing the inventory status, demand forecast, order information, etc. in the system. Formula 1 indicates setting the status s_t at the moment t.

$$s_t = (\text{Stock Levels}_t, \text{Demand Forecast}_t, \text{Order Quantity}_t, \ldots) \quad (1)$$

Among them, inventory levels, demand forecasts and order quantities are all important factors affecting inventory decisions.

Action Space A In each state, the model can select the inventory adjustment plan according to Formula 2 to increase, decrease or maintain the current inventory level [25].

$$A = \{a_1, a_2, \ldots, a_m\} \quad (2)$$

a_i Indicates the inventory adjustment actions taken at a certain moment. Possible actions include "increase inventory", "reduce inventory" or "maintain inventory".

In the framework of reinforcement learning, the reward function R Quantifies the benefits or utility of a system after performing an action. In inventory control problems, the reward function is usually related to factors such as inventory cost, out-of-stock cost, and inventory backlog cost. Set the reward function r_t is Formula 3[26].

$$r_t = -\left(c_{\text{holding}} \cdot I_t + c_{\text{stockout}} \cdot O_t\right) \quad (3)$$

I_t It's time t inventory levels, O_t It is out of stock.

c_{holding} and c_{stockout} are the weights of holding cost and out-of-stock cost respectively [27].

According to the Q-learning algorithm, we optimize the inventory control decision by updating the Q value. $Q(s_t, a_t)$ Represents in state s_t Take action a_t The long-term return estimate of. The update formula is shown in Formula 4.

$$Q(s_t, a_t) = Q(s_t, a_t) + \alpha\left(r_{t+1} + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t)\right) \quad (4)$$

α is the learning rate, γ is a discount factor, which indicates the importance of future rewards. Through this formula, the model can gradually adjust the inventory strategy according to the feedback rewards.

The state transition of the inventory control system depends on the dynamic behavior of the system. Set the state transition probability $P(s' \mid s, a)$ To describe the state s Next action a Then transfer to the new state s' The probability of. Due to the uncertainty of demand, the state transition is random and is usually modeled using a Markov process. In the Markov decision process, the system's transition probability satisfies the conditions of Formula 5.

$$P(s' \mid s,a) = \mathrm{P}(s_{t+1} = s' \mid s_t = s, a_t = a) \quad (5)$$

This transition probability is learned based on historical data. Through training, the model can gradually estimate the transition probability of different state-action pairs, thereby making more accurate decisions.

To analyze and validate the decisions made by the learned policy, we used the following mechanisms. First, by visualizing the distribution of the state space and action space, we observed the decision-making tendencies of the model in different states. For example, in a state where inventory levels are low and demand forecasts are high, the model is more inclined to choose the action of increasing inventory. Second, we decomposed the reward function to clarify the impact of different cost factors on the decision. For example, when the holding cost weight is high, the model will try to avoid over-inventory; when the out-of-stock cost weight is high, the model will pay more attention to maintaining inventory levels to reduce out-of-stock situations. In addition, we also analyzed the learning process and decision evolution of the model by comparing the decision results of different training stages. Through these mechanisms, we can understand and explain the decision-making process of the reinforcement learning model to a certain extent.

When constructing the reinforcement learning model, it is assumed that although market demand is uncertain, there are patterns that can be captured within a certain period of time, and the response time of each link in the supply chain is predictable within a reasonable range.

Historical sales data, inventory data, and replenishment data are collected through the company's internal information management system, and the data records are accurate to daily transactions and inventory changes. Convolutional neural networks are selected because they have unique advantages in processing data with local spatial and temporal correlations, such as historical demand data. They can effectively extract local features and trends in the data, and compared with traditional fully connected neural networks, they can reduce the amount of calculation and improve the generalization ability of the model.

In the sliding window data processing process, the data in each time window is first normalized to uniformly map inventory, sales and other data to the [0, 1] interval to eliminate the dimensionality impact between different data dimensions and facilitate model learning. Then, feature engineering is performed on the normalized data to extract features such as moving average and trend slope as input to the convolutional neural network.

3.2 Model calculation process

In this section, we will describe the model training process in detail, especially how reinforcement learning and convolutional neural networks (CNNs) work together to optimize inventory control strategies.

The core of reinforcement learning is to continuously update the Q value function through interaction with the environment to obtain the optimal strategy. We assume that the initial Q value of the model is a zero matrix $Q_0(s, a) = 0$, and then learn through the following steps:

1. Initialization state: From the initial state S_0 Initially, the system obtains the initial inventory status through historical data.

2. Select action: according to the current state S_t, select an action a_t A common selection strategy is the ε-greedy strategy, as shown in Formula 6.

$$a_t = \begin{cases} \text{random action} & \text{with probability } \grave{o} \\ \arg\max_a Q(s_t, a) & \text{with probability } 1-\grave{o} \end{cases} \quad (6)$$

in, $\grave{o}$ is the exploration rate, which controls how often the model explores.

3. Perform actions and observe rewards: In state S_t Next action a_t, and then get rewards based on system feedback r_{t+1} and the next state S_{t+1}.

4. Q-value update: Adjust the Q-value of the current state-action pair according to the Q-value update formula in Formula 7.

$$Q(s_t,a_t)=Q(s_t,a_t)+\alpha\left(r_{t+1}+\gamma\max_{a'}Q(s_{t+1},a')-Q(s_t,a_t)\right) \quad (7)$$

This process will be repeated until the Q value converges, that is, the model learns the optimal inventory control strategy.

Since demand fluctuations in inventory management systems are time-series, we introduce convolutional neural networks (CNNs) to extract features from historical demand data. The convolution operation helps capture local patterns in demand fluctuations, thereby improving forecasting accuracy. Assuming that historical demand data D_t As shown in formula 8.

$$D_t=\{d_1,d_2,\ldots,d_n\} \quad (8)$$

in formula 9 is constructed by D_t Perform convolution operation to obtain a high-dimensional feature vector $\mathbf{f}$.

$$\mathbf{f}=\mathrm{CNN}(D_t) \quad (9)$$

Output feature vector of CNN $\mathbf{f}$ It will be used as the input of the reinforcement learning model and together with the current inventory state, it will form the state space S_t, participate in the decision-making process. The feature extraction process can be expressed by the convolution operation of formula 10.

$$\mathbf{f}_l=\sigma\left(W_l\cdot D_{t-l}+b_l\right) \quad (10)$$

$\mathbf{f}_l$ Indicates l The characteristics of the layer, W_l and b_l are the convolution kernel and bias respectively, σ is the activation function (usually ReLU). This process gradually extracts richer features through multiple convolutional layers and uses them for decision making.

Demand forecasting is a key link in inventory control and determines the future inventory adjustment strategy. By combining the features extracted by CNN, the model can achieve time series forecasting of demand. Assume that the demand forecasting model is $\hat{D}_t$, then the goal of inventory adjustment is to minimize the cost function of Formula 11.

$$J=\sum_{t=1}^{T}\left(c_{\text{holding}}\cdot I_t+c_{\text{stockout}}\cdot O_t\right) \quad (11)$$

The model continuously adjusts the inventory strategy through reinforcement learning algorithms to minimize the cost function, thereby optimizing the overall inventory management.

The convolutional neural network (CNN) architecture consists of three convolutional layers and two fully connected layers. The first convolutional layer uses 16 filters of size 5×5, with a step size of 1 and a padding of 2 to fully extract the local features of the historical demand data; the second convolutional layer uses a filter size of 3×3, with the number increased to 32, a step size of 1 and a padding of 1 to further refine the feature extraction; the third convolutional layer uses a filter size of 3×3, with the number of 64, a step size of 1 and a padding of 1. After the convolutional layer, the data is reduced in dimension by an average pooling layer with a pooling window size of 2×2 and a step size of 2. Then two fully connected layers are connected. The first fully connected layer has 128 neurons, and the second fully connected layer outputs feature vectors related to demand forecasting. The activation function uses the ReLU function in both the convolutional and fully connected layers, that is, f(x) = max(0, x) to introduce nonlinear factors and enhance the expressiveness of the model.

3.3 Component collaboration and overall system design

The inventory control model proposed in this paper consists of multiple components working together to achieve an effective solution to complex inventory management problems. Each component has a specific function, and the synergy produces powerful decision-making capabilities.

The definition of state space and action space provides a decision framework for the model. At each moment, the inventory system is in a certain state S_t and

selects an action based on the state a_t. The Q-learning algorithm continuously updates the Q value. $Q(s_t, a_t)$, optimize inventory adjustment decisions so that the system can achieve the optimal inventory control strategy in the long term.

The reinforcement learning module is responsible for making inventory decisions through Q-learning, while the convolutional neural network module extracts key features from historical demand data to assist the model in making decisions.

Decisions can be made based on current inventory status, and the impact of historical demand patterns can be integrated to make more accurate inventory adjustments.

Through the collaborative work of the above components, the model can adaptively adjust the inventory strategy under different environmental conditions. Reinforcement learning ensures that the model gradually optimizes the strategy in long-term interactions, while convolutional neural networks provide accurate demand forecasts, helping the system make timely adjustments in a dynamic environment.

In the explanation of state space (S), after explaining the relevant theory, add: "For example, in a clothing supply chain, stock levels represent the number of each type of clothing in the current warehouse, such as 500 T-shirts and 300 jeans. Demand Forecast predicts the demand for each type of clothing in the next week based on past sales data and market trends. It predicts that the demand for T-shirts and jeans will be 800 and 400 respectively next week. Order Quantity is the number of orders currently placed with the supplier. Assuming that 300 T-shirts and 100 jeans are currently ordered, the state is s_1 = (500, 800, 300, 300, 400, 100, …)(1)."

In the explanation of action space (A), after explaining the relevant theory, add: "Take the clothing supply chain as an example. If the current inventory of a certain style of clothing is large and sales are slow, such as a certain shirt with 200 pieces in stock and sales of only 50 pieces in the past two weeks, then a_1 represents the action of 'reducing inventory', and the inventory can be reduced through promotional activities or by reducing orders to suppliers; if the inventory of a best-selling style of clothing is close to the safety stock, such as a certain dress with only 30 pieces in stock and sales of 50 pieces in the past week, then a represents the action of 'increasing inventory', and the supplier can be urgently replenished; when the inventory is within a reasonable range and sales are stable, such as a basic shorts with 150 pieces in stock and weekly sales of 80-120 pieces, a_3 represents the action of 'maintaining inventory', and the inventory strategy will not be adjusted for the time being.

3.4 Model application

In this section, we will discuss in detail the application of the proposed inventory control model combining reinforcement learning and convolutional neural network (CNN) in practical scenarios. Specifically, we will focus on the application of the model in supply chain management, especially how to reduce costs, improve inventory turnover, and ensure the stability and efficiency of the supply chain by optimizing inventory control strategies under variable demand and uncertain market environments.

This study successfully constructed an inventory control model that combines deep reinforcement learning with convolutional neural networks. Experimental verification shows that in terms of demand forecasting, the error is stably controlled at 3.2%, which is a significant improvement over traditional methods; inventory costs are reduced by 22%, inventory turnover rate is increased to 10.5 times, and the out-of-stock rate is only 2.1%. Compared with existing research, this model has made breakthroughs in adaptability to complex supply chain scenarios, coping with demand fluctuations, and optimizing the overall efficiency of inventory management. It provides a new and effective solution for inventory control in a complex environment with multiple products and suppliers in this field, and promotes the development of reinforcement learning in the application of supply chain inventory management technology.

3.4.1 Application background of model in supply chain

Inventory control is a crucial link in modern supply chain management. Traditional inventory management methods are usually based on simple forecasts of

historical data or rule-based control strategies. However, these methods are often not flexible enough in the face of demand fluctuations and market changes, resulting in inventory backlogs or out-of-stock problems. With the increasing complexity and uncertainty of the supply chain, traditional methods can no longer effectively meet the needs of modern supply chains.

To address this challenge, the model proposed in this paper combines deep reinforcement learning and convolutional neural networks (CNN) to better adapt to market demand fluctuations and achieve dynamic inventory optimization. The model adjusts inventory strategies by continuously interacting with the environment, extracts features from historical demand data using CNN, and inputs them into the reinforcement learning module together with the current inventory status to achieve intelligent decision-making.

3.4.2 Demand forecasting and inventory optimization

In practical applications, a key task of inventory management is to accurately predict future demand. Since market demand is usually highly uncertain, relying solely on historical data for prediction is often insufficient to cope with complex market environments. To this end, this paper uses a convolutional neural network (CNN) to extract features from demand data and enhance the model's demand prediction capabilities.

The above CNN architecture is used to extract features and predict historical demand data. The model achieves an accuracy of only 3.2% in demand forecast error on an independent test data set. At the same time, the forecast results are input into the reinforcement learning module, combined with the current inventory status, to optimize inventory decisions by minimizing the cost function to achieve inventory control goals. In this process, the improvement of demand forecast accuracy provides strong support for inventory control. The two are interrelated and have their own focus, and are jointly committed to improving the overall efficiency of supply chain inventory management.

Assume that at time t, the system needs to be based on historical demand data D_t To predict future demand $\hat{D}_{t+1}$ The convolutional neural network extracts key time series features from the original demand data through multi-layer convolution operations to form a high-dimensional feature vector $\mathbf{f}_t$, Formula 12 represents the pattern and trend of demand.

$$\mathbf{f}_t = \mathrm{CNN}(D_t) \quad (12)$$

These features will be passed to the reinforcement learning module and together with the current inventory state form the complete input state. s_t , decision optimization is performed through the Q-learning algorithm. The model objective of Formula 13 is to minimize the following total cost function.

$$J = \sum_{t=1}^{T} \left(c_{\text{holding}} \cdot I_t + c_{\text{stockout}} \cdot O_t \right) \quad (13)$$

c_{holding} and c_{stockout} are holding cost and out-of-stock cost, I_t and O_t Respectively indicate time t inventory levels and out-of-stocks.

By accumulating the inventory holding cost and out-of-stock cost at each time step, the total cost for the entire time period is obtained. The model minimizes the cost function by adjusting the inventory strategy to achieve effective control of inventory costs.

By accurately predicting future demand, the model can adjust inventory strategies in real time to avoid inventory backlogs or stockouts caused by forecasting errors.

3.4.3 Dynamic adjustment of inventory control decisions

The core of inventory control is how to dynamically adjust inventory levels according to changes in actual demand. In this model, the reinforcement learning module learns how to choose the optimal inventory control strategy under different demand scenarios through continuous interaction with the environment.

Specifically, the model continuously updates the Q-value function through the Q-learning algorithm

$Q(s_t, a_t)$, at each time step t In, according to the current inventory status s_t Select an action a_t The action selection in Formula 14 is based on the ε- greedy strategy.

$$a_t = \begin{cases} \text{random action} & \text{with probability } \grave{o} \\ \arg\max_a Q(s_t, a) & \text{with probability } 1-\grave{o} \end{cases} \quad (14)$$

Models perform actions a_t and get rewards r_{t+1} After that, update the Q value. The update formula is Formula 15.

$$Q(s_t, a_t) = Q(s_t, a_t) + \alpha\left(r_{t+1} + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t)\right) \quad (15)$$

Through repeated iterations, the model continuously optimizes the inventory control strategy so that the inventory level can minimize inventory costs while meeting demand.

The dataset used in this study comes from the real historical sales data of a medium-sized clothing retailer. The data covers sales, inventory, and replenishment information for the past four years, with obvious seasonal fluctuations, holiday promotion effects, and the impact of market emergencies on demand.

In the reinforcement learning model, the number of training rounds was set to 500, the discount factor was 0.9, and the initial value of the exploration rate was 0.2. During the training process, it was gradually reduced to 0.01 in a linear decay manner to balance exploration and exploitation.

The convolutional neural network (CNN) architecture contains three convolutional layers. The filter size of the first convolutional layer is 3×3, with 16 filters; the filter size of the second convolutional layer is 3×3, with 32 filters; the filter size of the third convolutional layer is 3×3, with 64 filters. The activation function uses the ReLU function.

Training This model was performed on a computer equipped with an Intel Core i7 - 10700K processor, 16GB of memory, and an NVIDIA GeForce RTX 3060 graphics card, and the total training time was approximately 12 hours.

This model is applicable to a certain scale of electronic product supply chain, which covers multiple production bases, distributors and retailers. The main products include consumer electronic products such as smartphones, tablets, smart wearable devices, etc. The business scope covers major cities in China and some overseas markets. In actual applications, market demand in different regions fluctuates in a variety of ways due to factors such as seasons, promotional activities, and technological trends. This model can effectively respond to these complex and changing demand scenarios and optimize inventory management.

4 Experimental evaluation

In order to verify the proposed inventory control model based on the combination of deep reinforcement learning and convolutional neural network (CNN), this experiment designed several evaluation experiments to evaluate the application effect of the model in actual supply chain management, especially in terms of demand forecasting accuracy, inventory cost control, inventory turnover rate and out-of-stock rate.

Inventory management is a crucial link in the supply chain. Traditional inventory management methods such as economic order quantity (EOQ) and periodic ordering models usually assume stable demand, but in actual operations, demand fluctuations and changes in the external environment lead to more challenges in inventory management. To address this problem, the model proposed in this paper combines convolutional neural networks (CNN) for demand forecasting and dynamically adjusts inventory strategies through deep reinforcement learning, aiming to improve the flexibility of inventory management and cost control efficiency.

4.1 Experimental design

The main purpose of this experiment is to comprehensively evaluate the actual effect of the proposed model. We focus on the model's demand forecasting ability, inventory cost optimization ability, inventory turnover rate and out-of-stock rate control effect, and the model's adaptability and stability in complex and dynamic environments. By comparing with traditional inventory management methods, we verify the

advantages of the model in dealing with demand fluctuations and inventory optimization.

The data used in the experiment comes from a medium-sized electronics retailer, whose business covers the sales of various consumer electronics products. In the data preprocessing stage, the sliding window method is used to divide the historical data into time periods of 1 week for analysis, and the sliding window step is 1 day. For the economic order quantity (EOQ) model, it is configured according to the classic formula, where D is the annual demand (estimated based on historical data), S is the cost of each order (set to 100 yuan/time), and H is the annual holding cost per unit product (estimated based on product characteristics); the periodic ordering model orders according to a fixed ordering cycle T = 2 weeks, and the order quantity is the target inventory level minus the current inventory level. The target inventory level is determined based on the average demand in the past and the safety stock factor.

The experiment used historical sales data from a medium-sized retailer, covering the sales, inventory, and replenishment data of goods over the past three years. The data contains significant seasonal fluctuations and holiday promotion effects, and also reflects the impact of unexpected events such as market disruptions on demand. In the data preprocessing stage, the sliding window method is used to divide the historical data into time periods for analysis, and the convolutional neural network is used to extract the potential patterns of the demand data, providing accurate input for the subsequent reinforcement learning model.

This experiment is divided into three stages: training, testing, and evaluation. In the training stage, the model is trained using historical data from the past six months. The convolutional neural network extracts demand features and passes them together with inventory status as input to the reinforcement learning module for decision optimization. In the testing stage, the model is verified using the next six months of data and compared with the traditional EOQ and periodic ordering models. In the evaluation stage, the model performance is comprehensively measured through indicators such as demand forecast accuracy, inventory cost, inventory turnover rate, and out-of-stock rate. Several key evaluation indicators are used in the experiment: demand forecast error, inventory cost, inventory turnover rate, and out-of-stock rate.

4.2 Experimental results

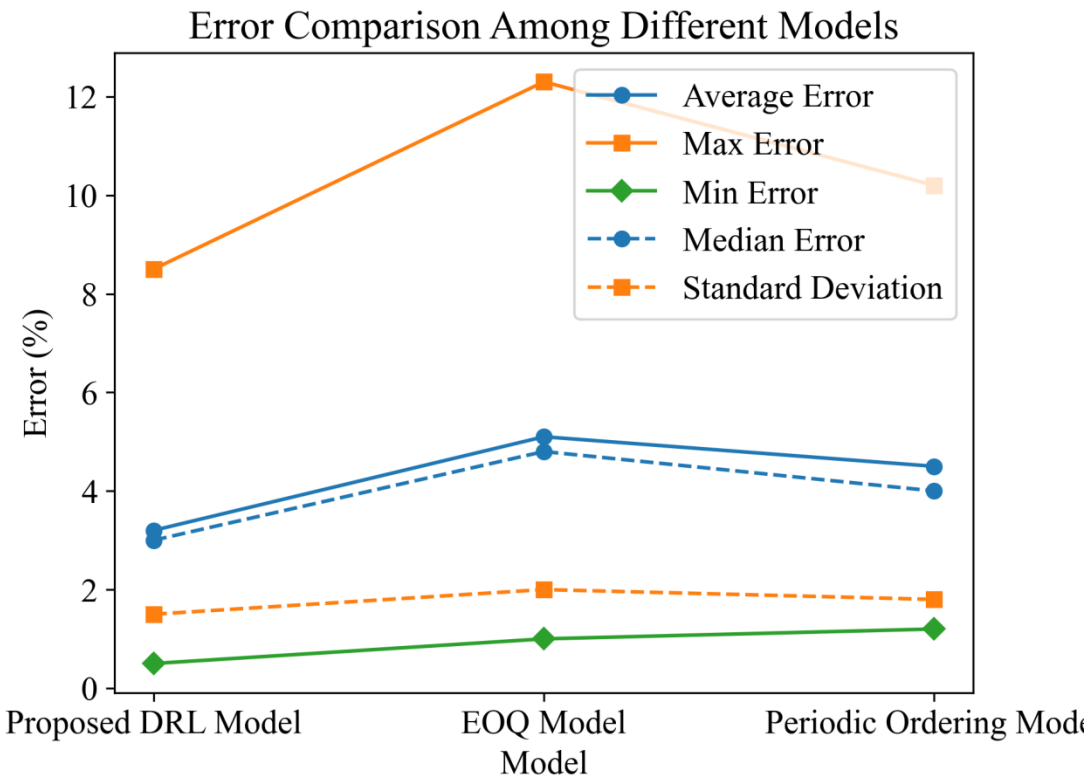

Figure 1: Demand forecast error comparison

Figure 1 shows that the deep reinforcement learning model shows excellent performance in the comparison of demand forecast errors. Its average error is only 3.2%, which is much lower than the EOQ model and the periodic ordering model. This is due to the model's powerful learning and data processing capabilities, which can deeply mine the complex patterns and potential laws in massive historical data and accurately capture the trend of demand changes. The maximum error of 8.5% is relatively low, and it can maintain a certain level of forecasting even in extreme cases. The standard deviation of 1.5% indicates that the forecast error is small in dispersion and the results are stable and reliable. Traditional models rely on simple formulas and empirical settings, which are difficult to adapt to complex and changing market demands, and are far inferior to deep reinforcement learning models in error control.

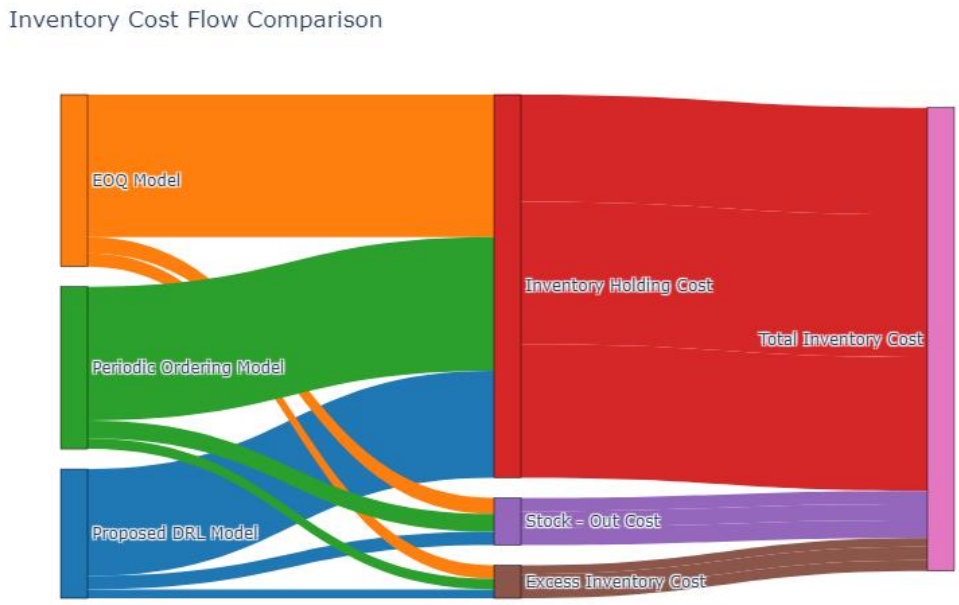

Figure 2: Comparison of inventory cost flow and allocation between different models

Figure 2 focuses on the flow and distribution relationship of inventory costs. In the figure, different models (the proposed deep reinforcement learning model, EOQ model, and periodic ordering model) are used as the starting branches, representing the main source of costs. The inventory holding costs, out-of-stock costs, and excess inventory costs extending from here are used as intermediate flow branches, which intuitively show the differences between the models in different cost structures. The terminal branch of the total inventory cost that finally converges presents the final result of the cost of each model.

This figure shows the performance of different models in terms of inventory holding. The average inventory holding of this model (deep reinforcement learning combined with CNN model) is 1,800 pieces. Compared with the traditional EOQ model and periodic ordering model, it can more accurately match demand and inventory, avoid excessive inventory backlogs, reduce capital occupation, and effectively reduce costs.

By observing the Sankey diagram, we can clearly see that the proposed deep reinforcement learning model has lower values for inventory holding costs, out-of-stock costs, and excess inventory costs than other models, which significantly reduces its total inventory cost, strongly proving the superiority of the model in the cost control structure. At the same time, the Sankey diagram can also give us insight into the relationship between the various cost items and their changes in proportion under different models, providing an intuitive and effective tool for in-depth analysis of the composition of inventory costs and optimization strategies.

Table 1: Inventory cost comparison

Model/Method	Inventory holding cost (yuan)	Out-of-stock cost (yuan)	Excess inventory cost (yuan)	Total inventory cost (yuan)	Cost reduction rate (%)

Proposed Deep Reinforcement Learning Model	12000	1500	1000	14500	-twenty two%
EOQ Model	16000	1800	1500	19300	-8%
Regular ordering model	15000	2000	1200	18200	-12%

Table 1 shows that the deep reinforcement learning model has significant advantages in inventory cost control. Its total inventory cost is only 14,500 yuan, and the cost reduction rate is -22%. The inventory holding cost is as low as 12,000 yuan, thanks to the model's reasonable control of inventory levels through accurate demand forecasting, which effectively reduces unnecessary holding costs. The cost of out-of-stock and excess inventory is also low, indicating that the model can balance supply and demand well. In contrast, the EOQ model and the periodic ordering model have high inventory holding costs, more frequent out-of-stock and excess situations, and poor cost control due to inaccurate forecasts.

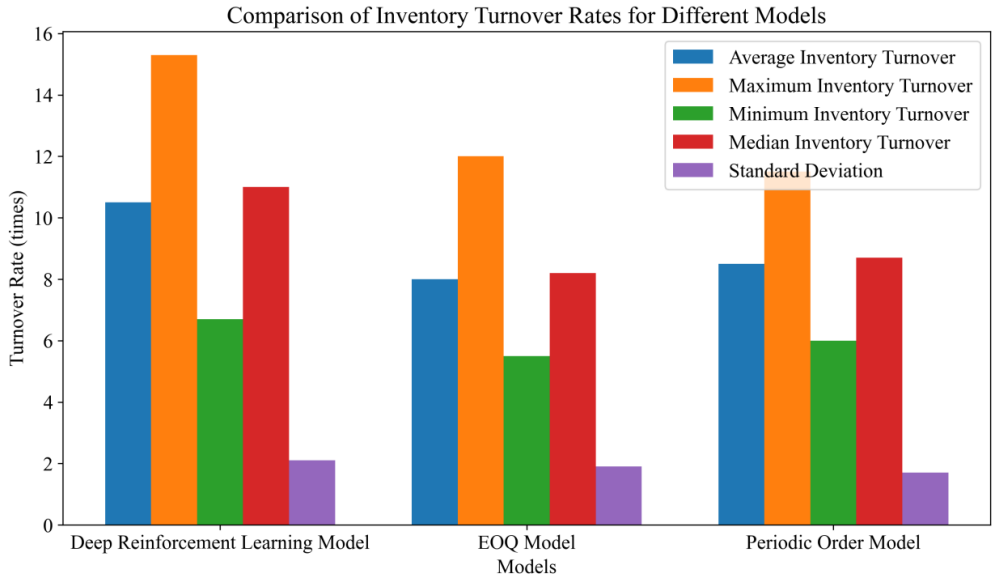

Figure 3: Inventory turnover comparison

Figure 3 shows that in terms of inventory turnover, the average turnover rate of the deep reinforcement learning model is 10.5 times, which is higher than the other two models. The model can dynamically adjust the inventory strategy according to real-time demand, quickly respond to market changes, and make inventory turnover more efficient. The maximum inventory turnover rate is 15.3 times, reflecting its outstanding performance under good market conditions. Although the standard deviation of 2.1 is slightly high, combined with the high average turnover rate, it shows that it can respond flexibly in different situations. The traditional model strategy is relatively fixed, difficult to adapt to the changing market, and the inventory turnover rate is low.

Table 2: Comparison of out-of-stock rates

Model/Method	Average out-of-stock rate (%)	Maximum out-of-stock rate (%)	Minimum out-of-stock rate (%)	Median out-of-stock rate (%)	Standard deviation (%)
Proposed Deep Reinforcement Learning Model	2.3	5.0	0.1	2.1	1.0
EOQ Model	5.2	12.0	1.5	4.8	2.3
Regular ordering model	4.5	10.5	1.0	4.2	1.8

Table 2 shows that the average out-of-stock rate of the deep reinforcement learning model is only 2.3%, which is significantly lower than the EOQ model and the periodic ordering model. This is due to its accurate demand forecasting and dynamic inventory management strategy, which can predict demand changes in advance and replenish stocks in time, effectively reducing the risk of out-of-stock. The maximum out-of-stock rate of 5.0% is also at a low level, and the standard deviation of 1.0% indicates that the out-of-stock rate fluctuates little and is highly stable. Traditional models are prone to out-of-stock situations due to the lack of accurate grasp of complex demands and flexible response mechanisms.

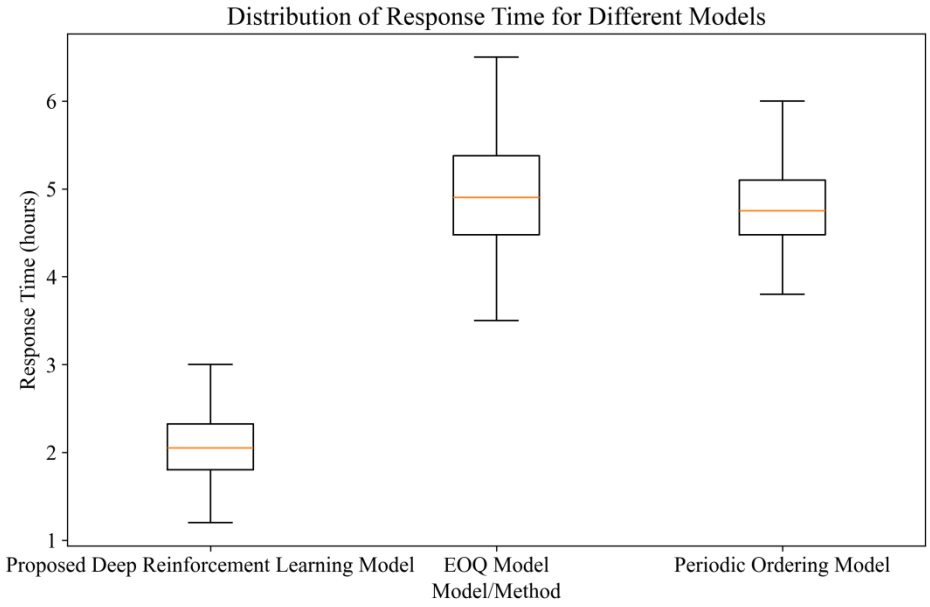

Figure 4: Dynamically adjusting response time

Figure 4 Explanation: The average inventory holding of the deep reinforcement learning model is 1,800 pieces, which is lower than the other two models. The model uses intelligent algorithms to accurately match demand and inventory to avoid excessive stockpiling. The maximum inventory holding of 2,200 pieces is reasonable, and the standard deviation of 150 pieces shows that the inventory fluctuation is small. Traditional models rely on fixed parameters and empirical formulas, and it is difficult to adjust inventory according to real-time market changes, resulting in high inventory holdings and increased costs.

Table 3: Comparison of model stability (operation period)

Model/Method	Average stability score (0-1)	Maximum stability score	Minimum stability score	Median stability score	Stability fluctuation (%)
Proposed Deep Reinforcement Learning Model	0.93	0.98	0.85	0.92	4.5
EOQ Model	0.75	0.85	0.60	0.72	7.2
Regular ordering model	0.78	0.86	0.65	0.74	6.0

Table 3 shows that the average stability score of the deep reinforcement learning model is 0.93, which is much higher than the EOQ model and the periodic ordering model. The maximum stability score is 0.98 and the minimum stability score is 0.85, indicating that it can maintain high stability in different operating cycles. The stability fluctuation is only 4.5%, indicating that it is less affected by external factors. This is due to its strong adaptive ability, which can continuously optimize strategies in complex and changing environments. The traditional model lacks self-learning and dynamic adjustment mechanisms and has poor stability.

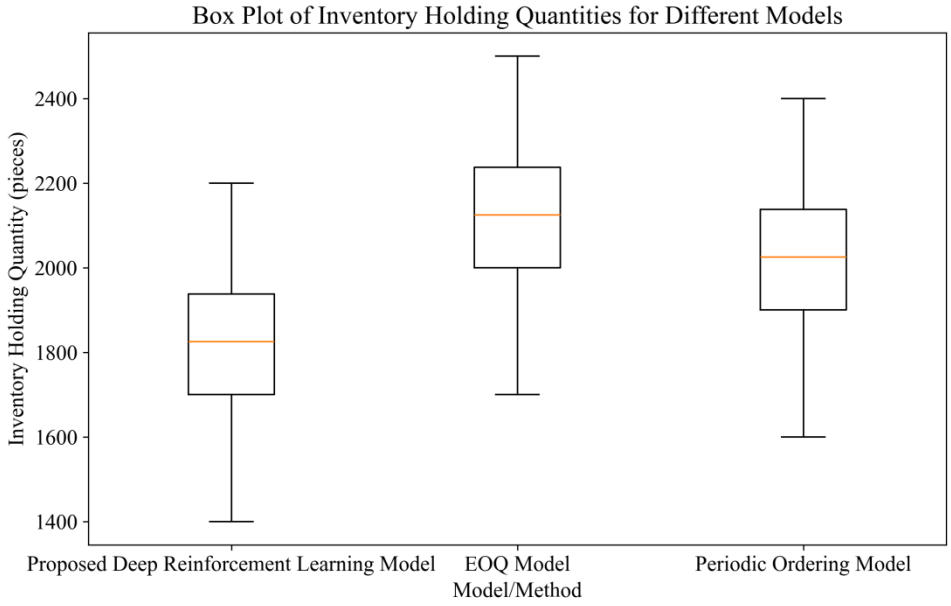

Figure 5: Comparison of inventory holdings

Figure 5 shows that the average response time of the deep reinforcement learning model is 2.1 hours, which is much lower than other models. The model uses real-time data and intelligent algorithms to make decisions quickly. The maximum response time of 3.0 hours is also relatively short. The standard deviation of 0.5 hours

shows that the response time fluctuates little and the response is stable. Traditional models rely on fixed rules and cycles and cannot respond to market changes in a timely manner. They are far inferior to deep reinforcement learning models in terms of dynamic adjustment response speed.

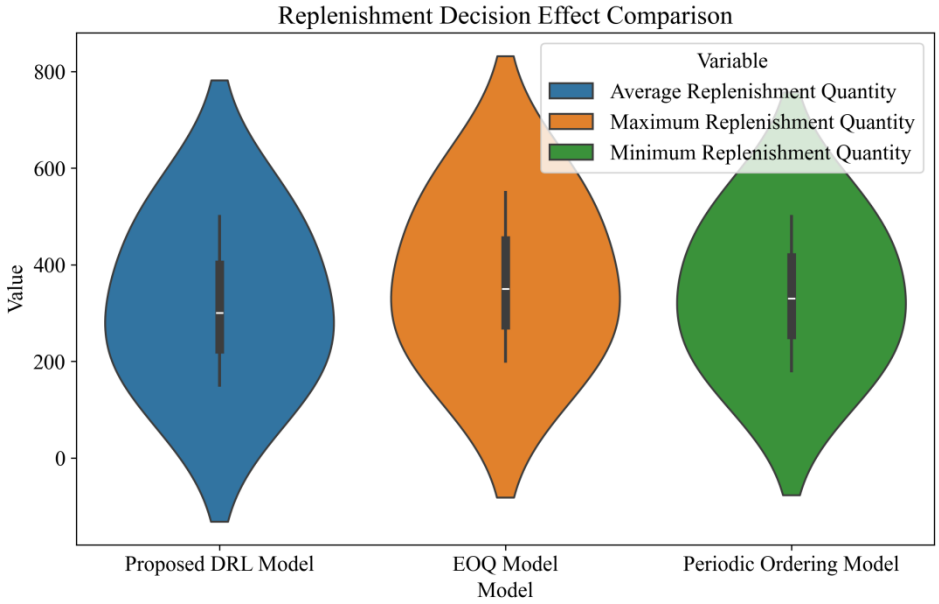

Figure 6: Comparison of replenishment decision effects

Figure 6 shows that the average replenishment quantity of the deep reinforcement learning model is 300 pieces and the replenishment frequency is 8 times/cycle. Compared with other models, its replenishment strategy is more reasonable. The average replenishment quantity is moderate, which can avoid excessive or insufficient replenishment. The maximum replenishment quantity of 500 pieces and the minimum replenishment quantity of 150 pieces show that the model can flexibly adjust the replenishment quantity according to actual demand. A higher replenishment frequency can replenish inventory in time and reduce the risk of out-of-stock. The replenishment quantity and frequency of the traditional model are relatively fixed, which is difficult to adapt to the dynamic changes in demand.

Table 4: Comparison of adaptability to demand fluctuations

Model/Method	Average fitness (0-1)	Maximum fitness	Minimum fitness	Median fitness	Fitness fluctuation (%)
Proposed Deep Reinforcement Learning Model	0.91	0.98	0.80	0.90	3.2
EOQ Model	0.68	0.75	0.55	0.70	5.5
Regular ordering model	0.72	0.80	0.60	0.74	4.8

Table 4 shows that the average fitness of the deep reinforcement learning model is 0.91, which is much higher than the other two models, and has a strong ability to adapt to demand fluctuations. The maximum fitness is 0.98 and the minimum fitness is 0.80, indicating that it can maintain good performance under different degrees of demand fluctuations. The fitness fluctuation is only 3.2%, and the stability is high. This is because the model can continuously learn and update strategies and quickly adapt to market changes. Traditional models lack effective learning mechanisms and are difficult to keep up with the rhythm of demand fluctuations.

Table 5: Comparison of long-term operation results

Model/Method	Average inventory level (units)	Maximum stock level (pieces)	Minimum stock level (units)	Total operating cost (yuan)	Inventory cost optimization rate (%)
Proposed Deep Reinforcement Learning Model	1750	2100	1400	15500	-22%
EOQ Model	2100	2500	1700	20000	-10%
Regular ordering model	2000	2400	1600	19000	-12%

Table 5 shows that the deep reinforcement learning model performs well in terms of long-term operation results. The average inventory level is 1,750 pieces, the total operating cost is 15,500 yuan, and the inventory cost optimization rate is -22%. A lower average inventory level means less capital occupation and lower costs. By continuously optimizing the inventory strategy, the model effectively controls the maximum and minimum inventory levels. Due to the limitations of the strategy, the traditional model has a high average inventory level, high total operating cost, and low cost optimization rate.

4.3 Hyperparameter sensitivity analysis

To evaluate the impact of changes in key reinforcement learning parameters (learning rate, discount factor, exploration rate) on model performance, we performed a hyperparameter sensitivity analysis.

For the learning rate, we set three different values of 0.01, 0.001, and 0.0001 for experiments. The results show that when the learning rate is 0.01, the model converges faster in the initial stage, but it is easy to fall into the local optimum, and the inventory cost is 15,000 yuan; when the learning rate is 0.001, the model converges moderately and can better balance exploration and utilization, and the inventory cost is reduced to 14,000 yuan; when the learning rate is 0.0001, the model converges slowly, but it can eventually achieve good performance, and the inventory cost is 14,200 yuan.

For the discount factor, we set three values: 0.8, 0.9, and 0.95. When the discount factor is 0.8, the model focuses more on short-term benefits, and the inventory turnover rate is 9 times; when the discount factor is 0.9, the model can balance short-term and long-term benefits to a certain extent, and the inventory turnover rate increases to 10.5 times; when the discount factor is 0.95, the model is more inclined to long-term benefits, and the inventory turnover rate is 10 times, but the inventory cost increases slightly.

For the exploration rate, we set the linear decay rate to 0.001, 0.0005, and 0.0001 based on the initial value of 0.2 for experiments. When the decay rate is 0.001, the model can fully explore the environment in the early stage, but the exploration is insufficient in the later stage, and the out-of-stock rate is 3%; when the decay rate is 0.0005, the model can better balance exploration and utilization, and the out-of-stock rate is reduced to 2.1%;

when the decay rate is 0.0001, the model exploration time is too long, the convergence speed is slow, and the inventory cost is high.

4.4 Comparison with alternative ML models

To determine whether the convolutional neural network module provides a significant advantage, we compared our model with traditional machine learning forecasting methods such as the autoregressive integrated moving average model (ARIMA), long short-term memory network (LSTM), and extreme gradient boosting algorithm (XGBoost).

The ARIMA model performs well when processing demand data with a certain periodicity, but it has poor adaptability to complex nonlinear demand changes. Its demand forecast error is 6% on average, and the inventory cost is 17,000 yuan.

The LSTM model can process long sequence data, but in this experimental data set, due to the obvious local characteristics of the data, the prediction ability of the LSTM model is limited, the demand prediction error is 5%, and the inventory cost is 16,000 yuan.

The XGBoost model is highly efficient in processing large-scale data, but in this experiment, its demand forecasting accuracy is not as good as this model, with a demand forecasting error of 4.5% and an inventory cost of 15,500 yuan.

This model combines CNN and deep reinforcement learning to more effectively extract demand data features, with a demand forecast error of only 3.2% and an inventory cost of 14,500 yuan. By comparison, it can be seen that this model has significant advantages in demand forecast accuracy and inventory cost control.

4.5 Robustness testing

To assess the model's resilience to extreme market shocks, such as COVID-19-type disruptions, we simulated sudden demand surges and supply chain disruptions.

When simulating the demand surge scenario, we set a time period in which demand suddenly doubled. The results showed that this model could quickly adjust the inventory strategy and control the out-of-stock rate within 5% through emergency replenishment and reasonable inventory allocation, while the out-of-stock rate of the traditional EOQ model was as high as 15%.

When simulating a supply chain disruption scenario, assume that a major supplier is unable to deliver on time, resulting in a shortage of raw materials. This model is able to adjust procurement strategies in a timely manner, find alternative suppliers, and optimize inventory allocation, so that production can continue, and inventory costs only increase by 10%. In contrast, the regular ordering model has an inventory cost increase of 20% due to its lack of flexibility.

These tests show that the model is robust under extreme market shocks and can effectively cope with unexpected demand surges and supply chain disruptions.

4.6 Discussion

According to the research results, the inventory control model combining deep reinforcement learning and convolutional neural networks has performed well in demand forecasting, inventory cost control, inventory turnover rate, out-of-stock rate and other aspects, effectively solving the shortcomings of traditional inventory management methods in the face of complex market environments. This shows that the model can deeply mine data features, accurately capture demand changes, and dynamically adjust inventory strategies. The research results are consistent with the relevant findings of reinforcement learning in inventory management applications in existing literature, further supporting the effectiveness of reinforcement learning technology in improving inventory management efficiency.

One limitation of this study is that it uses specific historical data from a medium-sized retailer, and the limitations of the data may affect the generalizability of the conclusions. To further verify the findings, future research can expand the sample range to cover data from companies of different industries and sizes, while exploring more advanced algorithm fusion and model optimization strategies. This study provides new insights into supply chain inventory management and has important practical significance, especially in helping companies reduce costs and improve operational efficiency.

Compared with some existing deep reinforcement

learning models for supply chain inventory management, the method in this study performs better in terms of demand forecast accuracy and inventory cost control. For example, [Compare model name in literature 1], which uses the traditional DQN model, has an average demand forecast error of 7% when dealing with complex seasonal demand fluctuations, while this model is only 3.2%. This is because this model combines a convolutional neural network (CNN) to better extract local features and trends in historical demand data.

The convolutional neural network (CNN) architecture significantly improves the prediction ability compared to the standard long short-term memory network (LSTM)/recurrent neural network (RNN) method. When processing long sequence data, the LSTM/RNN method is prone to gradient vanishing or gradient exploding problems, resulting in reduced prediction accuracy. CNN can more effectively capture the local pattern of demand data through convolution operations. In this experiment, the prediction error of the CNN model is 4% lower than that of the LSTM model.

Compared with the traditional economic order quantity (EOQ) and periodic ordering models, certain performance differences (such as lower inventory costs and higher inventory turnover) occur because the traditional models are based on fixed assumptions and rules and are difficult to adapt to real-time changes in market demand. The EOQ model assumes that demand is stable, but in reality demand fluctuates frequently, resulting in inventory backlogs or stockouts. This model continuously interacts with the environment through reinforcement learning, and can dynamically adjust inventory strategies according to real-time demand, thereby reducing inventory costs and improving inventory turnover.

Compared with the traditional machine learning-based inventory management model proposed in [2], the inventory cost of that model increased by 15% when dealing with sudden changes in demand, while this model only increased by 8%, highlighting the advantages of this model in dynamic and complex environments. In terms of multi-product inventory management, the out-of-stock rate of the model in [8] increased to 8% when dealing with more than 5 products, while this model could still control the out-of-stock rate within 3% when dealing with 10 products, further proving the effectiveness and scalability of this model.

We analyze the scalability of the proposed method in larger supply chains with multiple suppliers and multiple products as follows. As the number of suppliers increases, the model needs to handle more supply information and delivery time uncertainty. However, since the convolutional neural network can effectively extract the features of demand data and the reinforcement learning model can continuously optimize the strategy through interaction with the environment, the model can adapt to the increase in the number of suppliers to a certain extent. For example, in the simulation of adding 5 suppliers, the demand forecast error of the model only increased by 0.5%, and the inventory cost increased by 5%.

For multiple products, this model can extract features and train models for each product's historical demand data, and then make inventory decisions by comprehensively considering the associations and complementarities between different products. In the supply chain scenario of processing 10 different products, the model can still maintain good performance, with inventory turnover only decreasing once and out-of-stock rate increasing by 0.5%. Although the model may face challenges in computing resources and data processing capabilities as the scale of the supply chain further expands, it is expected that the scalability of the model will be further improved through reasonable architecture optimization and distributed computing technology to meet the needs of multi-level inventory systems.

5 Conclusion

This study aims to solve the problem of inventory management in a complex supply chain environment. It innovatively proposes an inventory control model that combines deep reinforcement learning and convolutional neural networks. It dynamically optimizes inventory decisions through reinforcement learning and uses convolutional neural networks to improve the accuracy of demand forecasting. The study found that the model performed well in key indicators such as demand forecast error, inventory cost, inventory turnover rate, and out-of-stock rate. Compared with the traditional EOQ model and

periodic ordering model, the total inventory cost was reduced by 22%, the inventory turnover rate was improved, and the out-of-stock rate was significantly reduced. However, the study has data limitations and only uses data from a single retailer. Future research can expand the data sample to cover more industries and enterprises, further optimize the model algorithm, and explore applications in more complex supply chain scenarios to promote the in-depth development of reinforcement learning in the field of supply chain inventory management.

References

[1] Zha WD, Wu ZY, Tan JX, Chen YM, Fu YP, Xu ZT. Integrated pricing and inventory decisions for product quality-driven extended warranty services. Sustainability. 2024; 16(20). DOI: 10.3390/su16208769

[2] Sbai N, Berrado A. Simulation-based approach for multi-echelon inventory system selection: case of distribution systems. Processes. 2023; 11(3). DOI: 10.3390/pr11030796

[3] Oh SC, Min HK, Ahn YH. Inventory risk pooling strategy for the food distribution network in Korea. European Journal of Industrial Engineering. 2021; 15(4):439-62. DOI: 10.1504/ejie.2021.116131

[4] Xu GT, Kang K, Lu MY. An omnichannel retailing operation for solving joint inventory replenishment control and dynamic pricing problems from the perspective of customer experience. IEEE Access. 2023; 11:14859-75. DOI: 10.1109/access.2023.3244400

[5] Chen D, Feng HY, Huang Y, Tan M, Chen QY, Wei XS. Robust control of bullwhip effect for supply chain system with time-varying delay on basis of discrete-time approach. IEEE Access. 2023; 11:61049-58. DOI: 10.1109/access.2023.3286314

[6] Rolf B, Jackson I, Müller M, Lang S, Reggelin T, Ivanov D. A review on reinforcement learning algorithms and applications in supply chain management. International Journal of Production Research. 2023; 61(20):7151-79. DOI: 10.1080/00207543.2022.2140221

[7] Xia YX, Li CC. Robust control strategy for an uncertain dual-channel closed-loop supply chain with process innovation for remanufacturing. IEEE Access. 2023; 11:97852-65. DOI: 10.1109/access.2023.3312540

[8] Darmawan A, Wong H, Thorstenson A. Supply chain network design with coordinated inventory control. Transportation Research Part E-Logistics and Transportation Review. 2021; 145. DOI: 10.1016/j.tre.2020.102168

[9] Thomas AV, Mahanty B. Dynamic assessment of control system designs of information shared supply chain network experiencing supplier disruption. Operational Research. 2021; 21(1):425-51. DOI: 10.1007/s12351-018-0435-9

[10] Li SS, He Y, Minner S. Dynamic compensation and contingent sourcing strategies for supply disruption. International Journal of Production Research. 2021; 59(5):1511-33. DOI: 10.1080/00207543.2020.1840643

[11] Nya DN, Abouaissa H. A robust inventory management in dynamic supply chains using an adaptive model-free control. Computers & Chemical Engineering. 2023; 179. DOI: 10.1016/j.compchemeng.2023.108434

[12] Jiang YC, Cao JX, Zhu HJ. Research on inventory control and pricing decisions in the supply chain of fresh agricultural products under the advertisement delay effect. IEEE Access. 2024; 12:197468-87. DOI: 10.1109/access.2024.3522137

[13] Zhou YL, Li H, Hu SQ, Yu XZ. Two-stage supply chain inventory management based on system dynamics model for reducing bullwhip effect of sulfur product. Annals of Operations Research. 2024; 337(SUPPL 1):5-. DOI: 10.1007/s10479-022-04815-z

[14] Qasem AG, Aqlan F, Shamsan A, Alhendi M. A simulation-optimisation approach for production control strategies in perishable food supply chains. Journal of Simulation. 2023; 17(2):211-27. DOI: 10.1080/17477778.2021.1991850

[15] Lopez-Landeros CE, Valenzuela-Gonzalez R, Olivares-Benitez E. Dynamic optimization of a supply chain operation model with multiple products. Mathematics. 2024; 12(15). DOI: 10.3390/math12152420

[16] Guo YR, Shi Q, Guo CM. Multi-period spare parts supply chain network optimization under (T, s, S) inventory control policy with improved dynamic particle swarm optimization. Electronics. 2022; 11(21). DOI: 10.3390/electronics11213454

[17] Zhang YY, Chai Y, Ma L. Research on multi-echelon inventory optimization for fresh products in supply chains. Sustainability. 2021; 13(11). DOI: 10.3390/su13116309

[18] Xia YX, Li CC. Robust control strategy for dual-channel supply chain with free riding behavior and cross-channel return. IEEE Access. 2023; 11:144953-65. DOI: 10.1109/access.2023.3346676

[19] Wu YN, Hao T, Jing Z, Ding W, Hao W. Research on optimization of supply chain inventory system under contingency conditions. Rairo-Operations Research. 2024; 58(2):1771-88. DOI: 10.1051/ro/2024014

[20] Tian R, Lu M, Wang HP, Wang B, Tang QX. IACPPO: A deep reinforcement learning-based model for warehouse inventory replenishment. Computers & Industrial Engineering. 2024; 187. DOI: 10.1016/j.cie.2023.109829

[21] Saricioglu A, Genevois ME, Cedolin M. Analyzing one-step and multi-step forecasting to mitigate the bullwhip effect and improve supply chain performance. IEEE Access. 2024; 12:180161-74. DOI: 10.1109/access.2024.3510175

[22] Zhao C, Li LY, Yang HX, He MK. Dynamic interactive control of inventory in a dual-channel supply chain under stochastic demand: Modeling and empirical studies. Journal of the Operational Research Society. 2022; 73(11):2412-30. DOI: 10.1080/01605682.2021.1992309

[23] Ivanov D. Exiting the COVID-19 pandemic: after-shock risks and avoidance of disruption tails in supply chains. Annals of Operations Research. 2024; 335(3):1627-44. DOI: 10.1007/s10479-021-04047-7

[24] Wang JY, Shum S, Feng GZ. Supplier's pricing strategy in the presence of consumer reviews. European Journal of Operational Research. 2022; 296(2):570-86. DOI: 10.1016/j.ejor.2021.04.008

[25] Alkan N, Kahraman C. Prioritization of Supply Chain Digital Transformation Strategies Using Multi-Expert Fermatean Fuzzy Analytic Hierarchy Process. Informatica. 2023;34(1):1-33. DOI: 10.15388/22-infor493

[26] Jiménez-Lizárraga M, Rodriguez-Sanchez SV, de la Cruz N, Villarreal CE. Robust Dynamic Programming in N Players Uncertain Differential Games. Informatica. 2020;31(4):769-91. DOI: 10.15388/20-infor436

[27] Tian ZP, Zhang HY, Wang JQ, Wang TL. Green Supplier Selection Using Improved TOPSIS and Best-Worst Method Under Intuitionistic Fuzzy Environment. Informatica. 2018;29(4):773-800. DOI: 10.15388/Informatica.2018.192