Hindawi
Mobile Information Systems
Volume 2021, Article ID 5880795, 11 pages
https://doi.org/10.1155/2021/5880795

Research Article

Application of Reinforcement Learning Algorithm in Delivery Order System under Supply Chain Environment

Haozhe Huang [1,2] **and Xin Tan** [3]

[1]*School of Economics and Management, Beijing Jiaotong University, Beijing 100044, China*
[2]*School of Economics and Management, Guangxi Vocational College of Performing Arts, Nanning 530000, Guangxi, China*
[3]*Computation Center of Guangxi, Jiaotong Investment Group of Guangxi, Nanning 530000, Guangxi, China*

Correspondence should be addressed to Xin Tan; 07244@aynu.edu.cn

Received 16 May 2021; Accepted 24 August 2021; Published 7 September 2021

Academic Editor: Sang-Bing Tsai

With the intensification of market competition and the development of market globalization, the efficiency of supply chain management orders has become an important part of enterprise competition resources. The competition among enterprises is fierce. To achieve effective customer response quickly, the time for supply chain order management is minimized, and refine the order processing process. This article introduces the strategy research of supply chain management order based on a reinforcement learning algorithm. This article first combines the reinforcement learning algorithm and deep learning algorithm, using the optimal decision-making ability of reinforcement learning algorithm and deep learning algorithm. The combination of data perception and the optimal ability to analyze examine the data of the order process, order cycle, and order delivery process of the supply chain order management and give the optimal decision. The supply chain order management process conducts questionnaire surveys and seminars to understand the current process of supply chain order management and the problems derived from the analysis of data based on the deep learning algorithm. Finally, through the output of the optimal strategy of the reinforcement learning algorithm, the supply chain order management process was improved, and the satisfaction survey was conducted again. The survey showed that the satisfaction was improved, and the satisfaction reached more than 90%.

1. Introduction

1.1. Research Background and Significance. With the rapid development of the times, the development of enterprises is also flourishing. The increase of enterprises and the innovation of supply chain management within enterprises have brought great challenges to enterprises [1]. Therefore, if an enterprise wants to develop continuously, it must continuously explore new supply chain management methods to improve its competitiveness. Among them, supply chain management is a new type of management mode and thought [2], and in supply chain management, in addition to the importance of suppliers, the management of supply chain orders is also essential, and the management of orders directly affects the entire enterprises; the management of orders involves customers at the source of purchase and the center of interest, so supply chain management has become an important method for enterprises to obtain sustainable competitive advantages [2, 3]. Supply chain management can shorten the order cycle, improve efficiency, and reduce the total cost of the enterprise. The main goal of inventory optimization is inventory-related expenses. The cost of capital increases with the increase in inventory. The backlog of goods is what companies should try to avoid. Reaching the benefits brought by supply chain management is the subject of concern for supply chain management strategy research [4–6]. Therefore, this article starts from the practice of the enterprise to discuss and study the problems of supply chain management and give reasonable opinions for the supply chain order management to achieve the benefits presented by the supply chain management.

1.2. Related Content. Supply chain management is increasingly regarded as the management of key business processes in the organizational network that makes up the supply chain [7, 8]. Croxton et al. exemplify the interface between processes and an example of how to implement process methods internally [9]. Because many people have realized the benefits of using process methods to manage the business and supply chain, most people are still unsure which processes to consider, which subprocesses and activities are included in each process, and how the processes interact with traditional functions isolated islands. Thus he believes that his goal in an organization is to provide managers with a framework for implementing supply chain management, provide lecturers with materials that can be used to build supply chain management courses, and provide researchers with a series of further development of the field opportunity [10, 11]. However, process management is only a key process and lacks a complete process. Gosavi proposed a reinforcement learning (RL) algorithm based on policy iteration [12] to solve average reward Markov and semi-Markov decision problems. His algorithm is an asynchronous, model-free algorithm (which can be used for large-scale problems). Its core idea is to calculate the value function of a given strategy and search in the strategy space [13, 14]. In the field of applied operations research, RL is used to provide good solutions to previously considered difficult problems. Therefore, he tested the proposed algorithm in commercial case studies related to practical problems in the aviation industry [15, 16]. In his experiments, he combined the algorithm with the nearest neighbor method to solve the larger state space [17, 18]. However, this kind of algorithm has a large error. Tao et al. proposed a steel manufacturing plant's production planning management system structure based on manufacturing to order (MTO) and manufacturing to inventory (MTS) management ideas [19]. In this architecture, he discussed the order planning process in detail and constructed a nonlinear integer programming model for the order planning problem. The model he proposed considers inventory matching and production planning at the same time and considers multiple objectives, such as the total cost of early/delay fines, delay fines within the delivery time window, fines for production, inventory matching, and cancellation orders, and he also considers the results of using PSO, TS, and hybrid PSO/TS algorithms to solve the models with three different orders compared [20, 21]. Numerical results show that the PSO/TS hybrid algorithm provides a better solution with high computational efficiency. However, while the calculation efficiency is improved, the accuracy of the calculation is uncertain.

1.3. Main Content and Innovation. The main content of this article is to study the strategy of supply chain order management based on reinforcement learning algorithms. Through questionnaire survey and discussion methods, we can obtain data on all aspects of supply chain order management in enterprises and carry out calculations on algorithms in reinforcement learning. The data is initially processed to obtain a series of problems in order management, and then through the combination of reinforcement learning algorithm and deep learning algorithm, problem analysis and processing and output of optimal strategies are performed, and then the problems in supply chain order management are given corresponding strategy recommendations. The innovation of this paper is to use the powerful data processing ability and optimal decision output ability of the reinforcement learning algorithm, combined with the data analysis of the algorithm and the accuracy of the optimal decision to analyze and give optimal strategy recommendations in the supply chain order management.

2. Concepts and Research Methods

2.1. Reinforcement Learning Algorithms and Supply Chain Management Concepts. Supply chain management integrates the functions of the enterprise with relevant data while improving the competitiveness of the enterprise. Supply chain management involves a wide range of issues [22]. In other words, it is a complex dynamic network structure that connects the material handling and channel selection of logistics management and coordination, pricing decisions, etc. With the widespread application of supply chain business models, how the ordering and pricing decisions of all members of the chain can satisfy everyone has become more important, and the more complex supply chain is composed of different enterprises; as the supply chain is composed of "chain," the transition to the "net" provides balanced decision-making changes in the process of gradual complexity in the supply chain structure [23, 24]. Supply chain management is also effective management of an enterprise, reflecting the strategic optimization of the whole process of an enterprise. According to the semi-Markov theory, the reinforcement learning algorithm is used to learn the joint supplementary problems in the supply chain without mentoring [25, 26]. In most cases, unpredictable emergency requirements will cause delays in delivery and reduce the efficiency of all links. In order to unite different supply chain links and solve these problems, the basic cycle of each kind of goods is used as the initial state. The Markov decision chain calculates the joint supplementary Q value through behavior and transition probability, parameter selection principles, and end conditions, and finally, the example verification proves the effectiveness and practicability of the algorithm [27, 28].

2.2. Reinforcement Learning and Deep Learning Combined Algorithm. Traditional reinforcement learning has a perfect theoretical model, and the algorithm is universal, but it has disadvantages such as low training efficiency and difficulty in processing high-dimensional data [29, 30]. Deep learning needs to go through complex network screening and physical linear transformation, can perform specific analysis of data, extract high-level representations of data, have powerful analysis capabilities for data, have a complete training mechanism, provide an approximate solution method for optimization problems, which can achieve the

best results in many applications [31]. Deep learning focuses on the analysis of data, and reinforcement learning has more advantages in the output of strategies [32]. Both algorithms have their own advantages. Deep reinforcement learning organically integrates the perception ability of deep learning and the decision-making ability of reinforcement learning. It can not only use deep learning to automatically learn information from large-scale input data but also uses reinforcement learning to make decision-making optimization based on this information. It is an end-to-end, end perception, and the control system has strong versatility. Therefore, there is an innovative deep reinforcement learning model [33].

2.3. Strategy Gradient Learning Algorithm. Strategy gradient enhancement learning PG-SVM multiround coordinated control method studies the joint supplementary problem of fuzzy variable demand under the condition of a single supplier; the demand is fuzzy variable; list its membership function; solve the objective function through trapezoidal fuzzy number; pass. The objective function is obtained by fuzzy membership degree, that is, the replenishment period of each product; the corresponding basic replenishment period length is determined by the optimal replenishment period of each product. Through the research on the joint supplementary problem of fuzzy demand, a reward function is obtained by the system after each action and the mathematical model is processed through the learning algorithm. The function finally solved is to minimize the order cost. Due to the large variance in the gradient estimation process, the convergence speed of the policy gradient algorithm is very slow, which has become an obstacle to the wide application of policy gradient reinforcement learning. At this time, it is assumed that the operation of the supply chain takes a week cycle, and a cycle consists of several systems. Competitive decision-making in the process of gradual complexity of the supply chain is composed of units. At this time, the cycle strategy is expressed as s. Without causing confusion, s will be denoted as si, and the goal of reinforcement learning is to find the optimal parameters to make the goal of reinforcement learning: the expectation of cumulative return is maximized. Supposing the state in the plot, the sequence of actions is consecutively arranged, R is the cumulative return of the plot, and p is the probability of the plot appearing under the strategy. At this time, the reported expectation can be expressed as follows:

$$x(\theta) = E[r|\pi]. \tag{1}$$

Similarly, in another action sequence, it can also be expressed as follows:

$$x(\theta) = \sum_{t} p(t|\theta)R(t). \tag{2}$$

Use the gradient method to optimize the objective function x:

$$\theta \leftarrow \theta + i\nabla x(\theta), \tag{3}$$

where i were the learning rate. Expand the gradient term in formula (3) to the following:

$$\nabla x(\theta) = \nabla_{\theta} \sum_{t} p(t|\theta)R(t). \tag{4}$$

Express x in the above formula with the state transition probability at each moment:

$$\nabla \log\ p(t|\theta) = \sum_{t=0}^{t-1} \nabla \log(x_t, a_t|\theta). \tag{5}$$

Combining the above two steps, we can get the following:

$$\nabla x(\theta) = \underset{t}{E}\left[\sum_{t=0}^{t-1} \nabla \log(x_t, a_t|\theta)R(t)\right]. \tag{6}$$

The expectation in formula (6) consists of two items. The first item is a direction vector; that is, the direction in which the probability of the current episode changes the fastest with the parameter t, and the parameter update in this direction can be increased or decreased to the greatest extent the probability of occurrence of plot t. The second term is a scalar, which plays a role in the degree of vector increase in the strategy gradient. The larger the R, the greater the vector increase. The intuitive meaning of the strategy gradient is to increase the probability of a high return trajectory and reduce the probability of a low return trajectory.

2.4. Questionnaire Survey Method. Select a company's order management related personnel to conduct a questionnaire survey. The survey content is divided into order process and existing problems, order delivery process problems, and a satisfaction survey after improvement. The investigation phase is divided into two parts. The first part is to investigate the company's existing order process and problems, collect data, and analyze, and the second part is to conduct a satisfaction survey on the improvements brought about by the previous part of the investigation. Obtain the satisfaction of the improved result.

2.5. Staff Discussion Method. The staff discussion method cannot be carried out directly. You must first formulate a detailed question sheet for the content of the information to be collected, grasp the problem raised the relationship between the order processing and the enterprise supply chain management insufficiency and purpose, and then collect it through face-to-face conversation information. The seminar method is not only to obtain information through questions but also to use conversations to actively guide more complete information. This article uses the staff discussion method to determine their dissatisfaction with the company's process and then uses the deficiencies they have reflected to improve the process.

3. Reinforcement Learning Problem and Supply Chain Structure Modeling

3.1. Modeling of Reinforcement Learning Problems. In the supply chain, the unit cycle cost of operators, distributors, and retailers is coordinated as the state model, and the order cycle time can also be the state model. When an enterprise order is placed, the main cost value of each cycle has been determined. The time and quantity of the order are both important factors. When the ordering time is too early, but the goods are stagnant in the warehouse, or the purchase of too many goods causes the goods to be backlogged in the warehouse, this will lead to an increase in inventory costs; too many orders cause excessive investment funds and long turnaround times, leading enterprises to increase the transportation cost of goods. As a result, the investment funds are too large and the turnaround time is too long, which leads to an increase in the cost of the enterprise. Too few orders need to be ordered as long as possible, which increases transportation costs. Good inventory management is to balance the question of when to order and how much to order. Therefore, inventory decision-making optimization has become an important link in the supply chain. The order cost increases with the increase in inventory. According to the basic economic order quantity, the total inventory cost in a period of time, the formula is as follows:

$$Tx(T,k,s) = C_h + C_p. \tag{7}$$

The latter equation is the inventory maintenance cost, which is a derivative of the surrounding costs of the inventory, such as the sum of the warehouse, water, and electricity costs, staff costs, site cost, insurance cost, etc. The long-term accumulation of goods will cause an increase in the latter equation. The size is shown in the following formula:

$$C_h = \frac{s}{t + \sum_{x=1}^{n} (s_x/k_x)} T. \tag{8}$$

Ordering costs are all costs in the process from the purchase of the goods to the warehousing after an order is placed, such as the public relations costs of the supplier, the transportation costs of the goods, the travel expenses of the purchaser, and so on. The order cost per unit period is shown in the following formula:

$$C_y = \sum_{i=1}^{n} D_i h_i \frac{T}{2}. \tag{9}$$

However, in direct life, the rate of demand for goods is often an uncertain factor; it is a fuzzy number. Therefore, combining the EOQ model and fuzzy numbers to establish a new fuzzy demand inventory model, the goal is still the lowest total inventory cost. The hypothesis is the demand rate n is a fuzzy variable with a known distribution, and the fuzzy variables are different in different periods; late delivery means that stocks are allowed to be out of stock during the supply process, but once they are out of stock, the owed goods must be replenished; all orders are delivered at one time; the demand per unit period is N; at this time, the maximum of the system is not the fuzzy demand, so the total system cost model is shown in the following formula:

$$TC(T,k,x) = \frac{\left(S + \sum_{i=1}^{n} (s/k)\right)}{T + \sum_{i=1}^{n} xkh(T/3)}. \tag{10}$$

According to fuzzy mathematics, x is expressed as a trapezoidal fuzzy number. Suppose the objective function:

$$f(x) = x, \quad x \in C. \tag{11}$$

Let M be the fuzzy maximum set of the formula on the fuzzy set, and its membership expression is shown in the following formula:

$$m(x) = \frac{f(x) - f(x)_{\max}}{f(x)_{\max} - f(x)_{\min}}. \tag{12}$$

The fuzzy decision set is shown in the following formula:

$$c_x = \begin{cases} \dfrac{x - m + n}{n - m + 2q} \\ \dfrac{-x + n + q}{q}, \\ 0 \end{cases} \tag{13}$$

where x is the conditional extreme point to be sought, the simplified fuzzy demand is shown in the following formula:

$$x = \frac{n(n - m + q) + q^2}{n - m - 2q}. \tag{14}$$

The optimal value of the end of the solution to the basic supplementary period is shown in the following formula:

$$T = \min\left(\sqrt{\left(S + \sum_{i=1}^{n} \frac{x}{k} \right), \frac{2}{xkh}} \right). \tag{15}$$

3.2. Supply Chain Model. The research in this paper mainly considers a three-level supply chain system that includes retailers, wholesalers, and manufacturers, and each level contains only one merchant. The three-level supply chain system is closely related to supply chain order.

As shown in Figure 1, the relationship at all levels of the supply chain is as follows. Retailers contact the market, accept consumer demand for goods in the market, predict, control inventory, and issue ordering requirements to the higher level after receiving the demand. Consumers buy goods from retailers, and retailers meet consumer demand by selling consumers a corresponding number of goods while inventory is reduced. The retailer adopts a certain ordering strategy to pass its demand for goods to the wholesaler by adopting a certain ordering strategy based on its own inventory reduction and combined with the forecast of future market development needs. After receiving the order, the wholesaler will make an ordering decision based

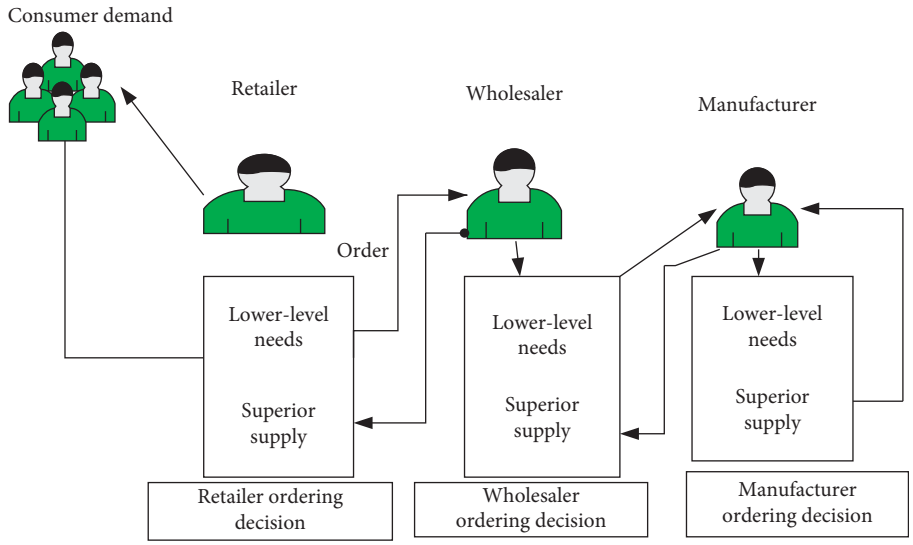

Figure 1: Supply chain model.

on the retailer's demand for the goods and the inventory situation after meeting the retailer's demand and send his demand for the goods to the manufacturer.

3.3. Reinforcement Learning Process. As shown in Figure 2, H represents the environment and s represents the initial state of the system. The learning process is as follows: First, the state sensor Z perceives the environment H and processes the information through the signal acquisition system to obtain the initial state s of the environment, and then the state sensor transforms s and sends a signal to the action selector D and the learner X, the selector D takes action b according to the learned knowledge and signal a and affects the environment. Because the environment H is affected by the agent's behavior, H changes. At this time, the environment variable is s. At the same time, the environment H feeds back a signal R to the template as a function of action b on the state. The learner X will change the strategy, and some feedback will come back. Strengthen the signal R and the internal signal a. It can be seen from the structure of the reinforcement learning system that the perception of state signals during the learning process plays an important role in the entire learning process.

3.4. Deep Reinforcement Learning Model. The depth enhancement model is shown in Figure 3. This model only uses the original video image information as input, after network processing, maps to the connection layer, and finally outputs the optimal value. This model has achieved results beyond the human level and has more advantages than traditional reinforcement learning algorithms in data input and analysis. Both of these two algorithms can output the optimal value, but the data analysis ability of deep reinforcement learning is relatively strong.

As shown in Figure 3, the deep reinforcement learning model is first input into the data. The data enters the network processing and then reaches the connection layer. The

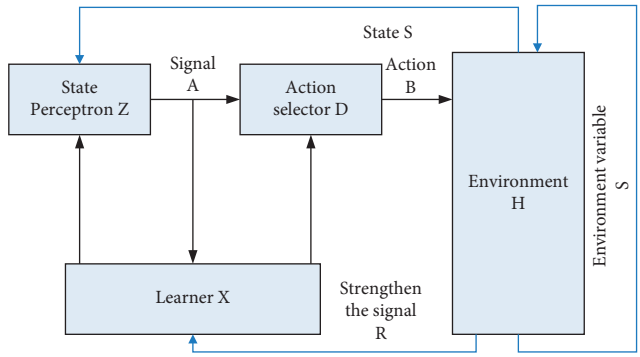

Figure 2: Reinforcement learning system structure.

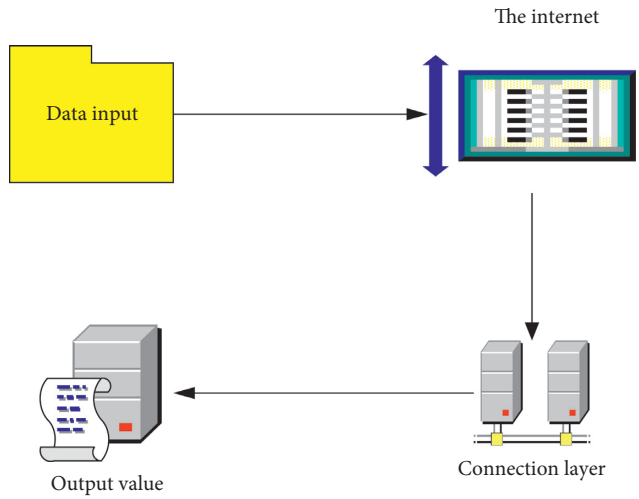

Figure 3: Deep reinforcement learning model.

connection layer analyzes again and finally obtains the output value. In this paper, the DQN network is introduced in the target detection task to learn the search strategy for candidate regions. The basic related theories of DQN will be explained below. A series of actions, observations, and

circulation will be carried out in the agent and the environment. At each time step, the agent selects an action from a set of actions. The action will be passed to the simulator, and its internal state and output score will change. Under normal circumstances, the environment is random, and the agent will not observe the real internal state of the simulator. What the agent sees from the simulator is only the original pixel array representing the current screen. During the interaction, the agent will receive changes in the representative data score from the simulator. Generally speaking, the score can depend on the complete sequence of previous actions and observations, and feedback on actions may take thousands of steps to reflect. Since the agent only observes the image on the current screen; that is to say, the information he observes is only a partial description of the internal state of the simulator; that is, it is impossible to fully understand the current state only from the current screen, therefore, the deep reinforcement learning model relies on a sequence of actions and observations to learn the strategy of the entire game.

4. Supply Chain Order Price and Process Analysis

4.1. Sensitivity Analysis of Order Inventory Price Name. According to the problem modeling of reinforcement learning, the relationship between purchase cost and the price is calculated numerically.

As shown in Table 1, when the cost is a change of 15–18, the price $P1$ rises from 31 to 39, and the price $P2$ also rises from 40 to 45. The decrease in cost and the increase in price lead to a decrease in profit and profit as unit costs increased and prices increased; the profits fell from 4000 to 3300. This phenomenon can be explained as that with the increase of purchase cost, the company's profit decreases compared to when the product purchase cost is low. To increase the company's profit, the company only has to increase the product's selling price $p1$ and $p2$. This shows that the changes of different parameters will have an unfavorable impact on the order quantity of the product and the price of different sales stages. Through this analysis, the understanding of product inventory management can be further strengthened, and the correct sales price and price can be set for the company—procurement strategy to help.

4.2. Analysis and Optimization of Order Shipping Process. The current delivery process is that the sales department issues a delivery plan instruction to the logistics system based on the delivery time required by the contract order. According to this plan, the logistics plans to communicate with the carrier on the delivery line after verifying the inventory quantity, and deploy the vehicle on the warehouse, and will load the goods and complete the shipment. If there is no suitable vehicle on the route that day, it will be transferred to the next day for delivery, and it is agreed that the delivery task must be completed within the next day. The specific process is as in Figure 4.

Table 1: Sensitivity analysis of parameter cost price.

	Cost	Price $p1$	Price $p2$	Profit
1	15	31	40	4000
2	16	34	41	3800
3	17	37	43	3500
4	18	39	45	3300

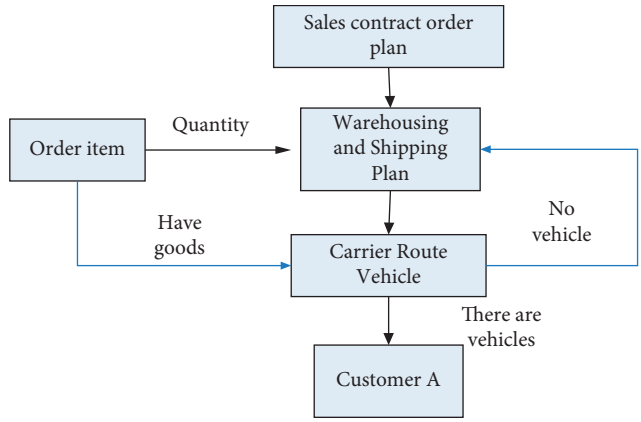

Figure 4: The original logistics delivery process transfer diagram.

It can be seen from Figure 4 that the ordered goods will be first planned for warehousing and delivery and will be sent directly to the customer when there are goods. When there is no vehicle, the vehicle cannot be shipped and can only be returned to the warehousing. The delivery instruction is issued by the sales department according to the time of the customer, A contract order, and the warehousing and delivery department only unilaterally execute it. If the carrier has the corresponding vehicle to the customer's location, the goods can be shipped. If there is no corresponding line, vehicles cannot be delivered even if there are goods in the warehouse. The main reason for the low turnover rate of warehouse goods with orders is the uncertainty of vehicle resources and the instability of warehousing and delivery plans.

In response to the problems in the above-mentioned process, we conducted on-site visits to the logistics delivery plan management and discussed the various routes to the customer's location with the carrier to determine the key to the problem. Through the on-site understanding of the site, the shipping plan and the sales plan are only one-way work, and the model is as follows:

As shown in Figure 5, after the analysis of the above process, it is found that the current information is only unilaterally transmitted, and it is only shipped according to the sales plan. If there are not enough vehicles, you can only wait, and it does not give full play to the carrier's advantages in gathering vehicle information and overall planning. The intermediate process is single and lacks integration and intercommunication. Therefore, the shipping process is redesigned as in Figure 6.

As shown in Figure 6, during the delivery process, the carrier actively provides the vehicle route information to the supply chain logistics delivery system. Logistics delivery no

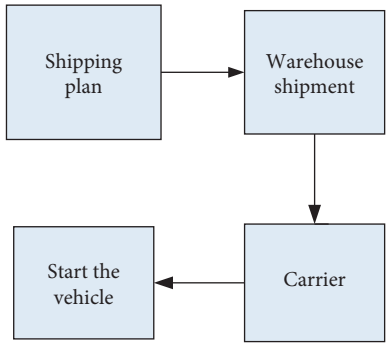

Figure 5: Eisting delivery process transfer diagram.

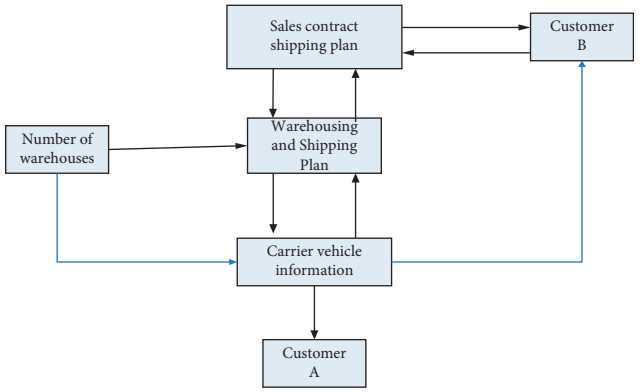

Figure 6: New shipping plan process.

Table 2: Comparison of warehouse turnover rate days.

Time	The first season	Second quarter	The third quater	Fourth quarter
The first half of the year	40	35	36	35
The second half of the year	30	25	26	25
Increase (%)	25	28	23	28

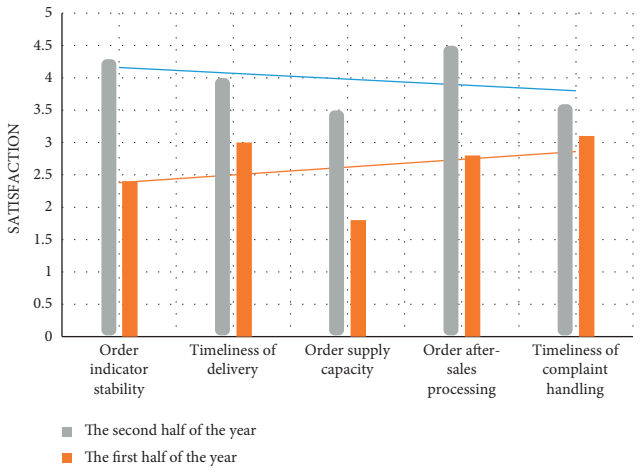

Figure 7: Customer satisfaction survey analysis diagram.

longer simply waits for the sales department's delivery plan but actively provides the vehicle information to the sales department. According to the vehicle information, the sales department negotiates and communicates with customers in the area on this route, signs contracts, and confirms the shipment of goods. These not only effectively use vehicle

Table 3: Problems in the order cycle and the results.

Problem	Result
Various customer order types and ordering methods	Orders are chaotic, resulting in missing orders
Unstable customer demand	When demand stops, production cannot be stopped, causing losses
Unstable customer demand	The procedures are complicated and complicated, resulting in a long-order cycle
Backward order processing methods	Long processing time
Unstable supply	Long procurement cycle

Table 4: Reasons and optimization of long-order cycle.

Reason	Optimize the target
Missing electronic order management system	Establish an electronic system to improve efficiency
Electronic material procurement lead time is too long and key materials are unstable	Procurement of key materials ahead of schedule
Complex delivery procedures	Simplify the delivery process
Customers choose whatever	Raise the threshold of cooperative customers, and select high-quality customers

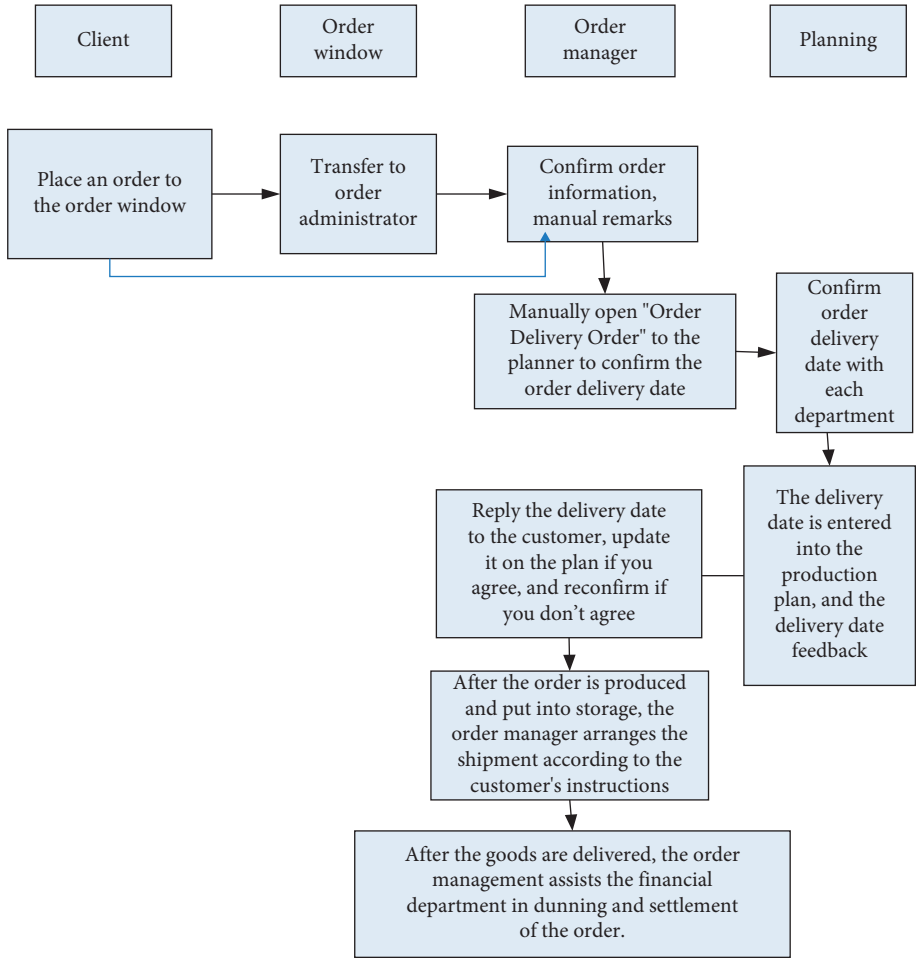

Figure 8: Flow chart of existing orders.

resources but also increase the overall processing and information management of vehicle information. And it objectively promotes sales activities and improves the turnover rate of goods.

As shown in Table 2, the turnover rate of each quarter has increased, and the number of turnover days has decreased. It increased 25% in the first quarter, 8% in the second quarter, 23% in the third quarter, and 28% in the fourth quarter. This

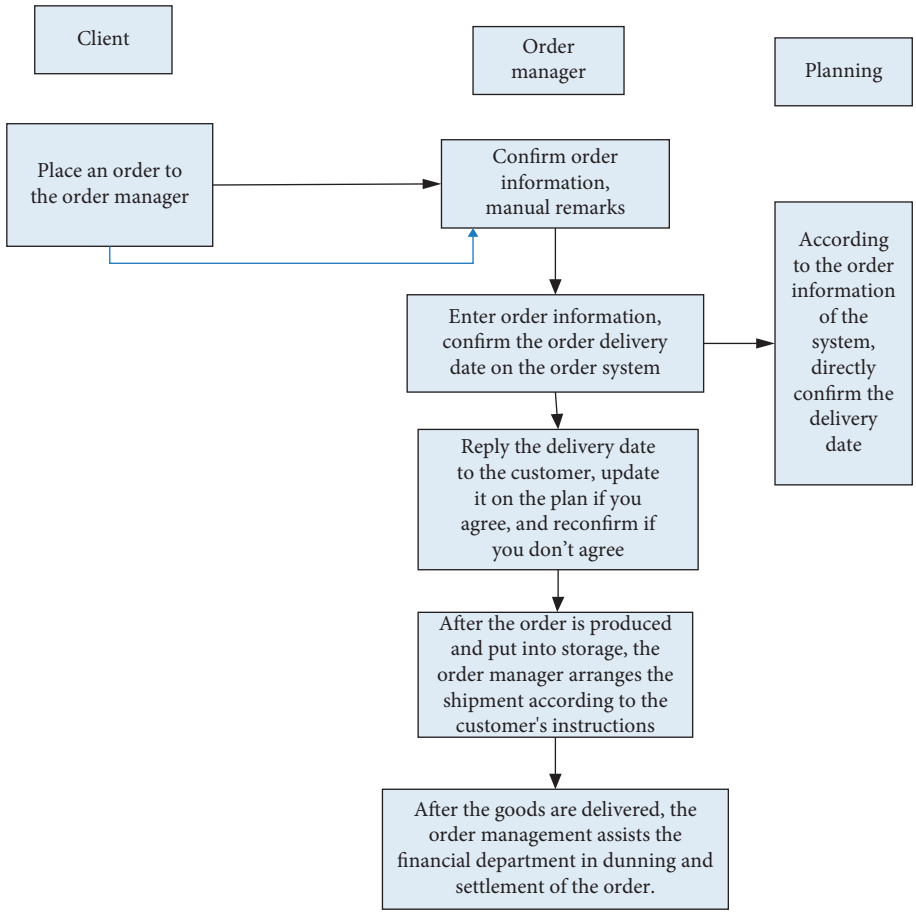

Figure 9: Flow chart improvement.

greatly reduces the amount of funds occupied by inventory materials, reduces management costs, and avoids the production of sluggish materials and the resulting losses and environmental protection issues. At the same time, after adopting the new order shipping process, the customer satisfaction and other delivery indicators are investigated, and the satisfaction survey data is analyzed as in Figure 7.

As shown in Figure 7, customer satisfaction in the second half of the year is higher than that in the first half of the year, especially in the stability of order indicators; the difference in satisfaction is the largest; the supply capacity of the order, the timeliness of delivery, the after-sales processing of the order, and the complaint satisfaction with the timeliness of processing have increased. It can be seen that the management of the new shipping process is effective for order management. Through the optimization of supply chain management, the optimization of order management has been realized, and the work efficiency has been greatly improved. The optimization of order processing and order delivery has shown good results.

4.3. Problems in Order Cycle and Optimization Analysis. The method of employee discussion and questionnaire survey was adopted for a company to analyze the information collected about the order cycle problems and found that the following problems mainly exist in the order fulfillment process.

As shown in Table 3, the existing problems include diverse customer order types and ordering methods, unstable customer demand, diverse order shipping modes, backward order processing methods, and unstable supply. It is easy to form order confusion and cause missed orders. Moreover, due to the complicated procedures, the order cycle is long, which will lead to the loss of customers. These are the main problems that cause the long-order cycle.

Therefore, in view of the various above-mentioned problems, the reasons and optimization goals are analyzed.

As shown in Table 4, the reasons for the occurrence are explained by the lack of an electronic processing system, long procurement, complex delivery process, and the pros and cons of cooperative customers, and optimization suggestions for the reasons are given. The first is to establish an electronic system for order processing and then advance the purchase of key materials to prevent material instability, shorten the purchase time, simplify the delivery process, and increase the threshold for cooperative customers, and select high-quality customers.

4.4. Optimization of Order Processing Process. The figure below is a general process of the first order.

As shown in Figure 8, there are many steps in the existing order process, and most of the orders are processed manually, and there are still too many departmental processing flows in the processing process. Therefore, the existing process wastes a lot of time, and this situation needs to be improved.

After being familiar with the order process and extensive suggestions from the company's senior management and related departments, the existing process was improved. The improvements are shown in Figure 9.

As shown in Figure 9, an order receiving window is canceled, orders are directly connected to customers by the order management office, and the previous manual operations are changed to electronic system operations. This not only improves the accuracy of order processing but also reduces order. The workload of the administrator saves processing time, and the order processing cycle time before and after the improvement is significantly shortened. It has fundamentally solved the missing order phenomenon caused by the instability of personnel, and the electronic operation has increased the order delivery time and the preorder processing time.

5. Discussion

This paper analyzes the supply chain order management based on the reinforcement learning theory, analyzes and improves the problems existing in the order processing process, the order transfer process, and the order cycle, and then combines the reinforcement learning algorithm for problem data processing and analysis and gives the optimal strategy. It improves the order processing and transshipment process, and at the same time makes strategic recommendations on the order cycle, uses an electronic system office, abandons the traditional manual processing of orders, improves the efficiency of order processing, simplifies the process, and shortens the order cycle, which is conducive to the improvement of the enterprise.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was partly supported by the fund coded, a National Natural Science Fund program (71831001).

References

[1] C. Rajkumar, L. Kavin, X. Luo, and J. Stentoft, "Doctoral dissertations in logistics and supply chain management: a review of Nordic contributions from 2009 to 2014," *Logistics Research*, vol. 9, no. 1, pp. 1–18, 2016.

[2] L. Lei, Z. Wang, and H. Zhang, "Adaptive fault-tolerant tracking control for MIMO discrete-time systems via reinforcement learning algorithm with less learning parameters," *IEEE Transactions on Automation Science and Engineering*, vol. 14, no. 1, pp. 299–313, 2017.

[3] H. Ying, Z. Zheng, F. R. Yu et al., "Deep reinforcement learning-based optimization for cache-enabled opportunistic interference alignment wireless networks," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 11, pp. 10433–10445, 2017.

[4] E. Duryea, M. Ganger, and H. Wei, "Exploring deep reinforcement learning with multi Q-learning," *Intelligent Control and Automation*, vol. 7, no. 4, pp. 129–144, 2016.

[5] X. Li, H. Jianmin, B. Hou, and P. Zhang, "Exploring the innovation modes and evolution of the cloud-based service using the activity theory on the basis of big data," *Cluster Computing*, vol. 21, no. 1, pp. 907–922, 2018.

[6] S.-B. Tsai, Y.-M. Wei, K.-Y. Chen, L. Xu, P. Du, and H.-C. Lee, "Evaluating green suppliers from green environmental perspective," *Environment and Planning B: Planning and Design*, vol. 43, no. 5, pp. 941–959, 2016.

[7] L. Liu, Z. Wang, and H. W. Zhang, "Adaptive fault-tolerant tracking control for MIMO discrete-time systems via reinforcement learning algorithm with less learning parameters," *IEEE Transactions on Automation Science and Engineering: A Publication of the IEEE Robotics and Automation Society*, vol. 14, no. 1, pp. 299–313, 2017.

[8] B. Steunebrink, P. Wang, and B. Goertzel, "Artificial general intelligence," in *Proceedings of International Conference on Artificial General Intelligence*, pp. 354–362, New York, NY, USA, July 2016.

[9] K. L. Croxton, S. J. García-Dastugue, D. M. Lambert, and D. S. Rogers, "The supply chain management processes," *International Journal of Logistics Management*, vol. 12, no. 2, pp. 13–36, 2016.

[10] W. Baumung and V. Fomin, "Framework for enabling order management process in a decentralized production network based on the blockchain-technology," *Procedia CIRP*, vol. 79, no. C, pp. 456–460, 2019.

[11] A. Aalaei and H. Davoudpour, "Two bounds for integrating the virtual dynamic cellular manufacturing problem into supply chain management," *Journal of Industrial and Management Optimization*, vol. 12, no. 3, pp. 907–930, 2017.

[12] A. Gosavi, "A reinforcement learning algorithm based on policy iteration for average reward: empirical results with yield management and convergence analysis," *Machine Learning*, vol. 55, no. 1, pp. 5–29, 2018.

[13] S. Aptea and N. Petrovsky, "Will blockchain technology revolutionize excipient supply chain management?," *The Journal of Excipients and Food Chemicals*, vol. 7, no. 3, pp. 76–78, 2016.

[14] P. Samaranayake, "A conceptual framework for supply chain management: a structural integration," *Supply Chain Management*, vol. 10, no. 1, pp. 47–59, 2017.

[15] A. M. Quarshie, A. Salmi, and R. Leuschner, "Sustainability and corporate social responsibility in supply chains: the state of research in supply chain management and business ethics journals," *Journal of Purchasing and Supply Management*, vol. 22, no. 2, pp. 82–97, 2016.

[16] D. Mishra, A. Gunasekaran, T. Papadopoulos, and S. J. Childe, "Big Data and supply chain management: a review and bibliometric analysis," *Annals of Operations Research*, vol. 270, no. 1, pp. 1–24, 2016.

[17] C. Piera, C. Roberto, and E. Emilio, "Environmental sustainability and energy-efficient supply chain management: a review of research trends and proposed guidelines," *Energies*, vol. 11, no. 2, p. 275, 2018.

[18] D. Fitrianah, T. Palito, and U. Salamah, "Implementation of ElasticSearch search engine on order management system data," *International Journal of Computer Applications*, vol. 181, no. 8, pp. 25–35, 2018.

[19] Z. Tao, Y. J. Zhang, Q. Zheng, and P. M. Pardalos, "A hybrid particle swarm optimization and tabu search algorithm for order planning problems of steel factory based on the make-to-stock and make-to-order management," *Journal of Industrial and Management Optimization*, vol. 7, no. 1, pp. 31–51, 2017.

[20] P. de Bie, R. Tepaske, A. Hoek, A. Sturk, and E. van Dongen-Lases, "Reduction in the number of reported laboratory results for an adult intensive care unit by effective order management and parameter selection on the blood gas analyzers," *Point of Care: The Journal of Near-Patient Testing & Technology*, vol. 15, no. 1, pp. 7–10, 2016.

[21] T. Pieper, A. Schröder, and A. Hoffjan, "Management accounting change am beispiel der Energiewirtschaft," *Zeitschrift für öffentliche und gemeinwirtschaftliche Unternehmen*, vol. 41, no. 4, pp. 322–336, 2018.

[22] H. Matsuno, T. Hata, H. Takahashi et al., "A safety management case of laparoscopic colectomy in a patient with paroxysmal nocturnal hemoglobinuria," *International Surgery*, vol. 104, no. 7-8, pp. 333–337, 2019.

[23] G. V. Ramesh, K. Palanna, M. Bharath, H. V. Kumar, and T. Nagaraja, "Assessing the in vitro efficacy of new molecules of fungicides against Bipolaris setariae infecting browntop millet," *Journal of Pharmacognosy and Phytochemistry*, vol. 10, no. 2, pp. 1123–1130, 2021.

[24] A. Tharwat, H. Mahdi, M. Elhoseny, and A. E. Hassanien, "Recognizing human activity in mobile crowdsensing environment using optimized k-NN algorithm," *Expert Systems with Applications*, vol. 107, pp. 32–44, 2018.

[25] A. Matzke, T. Volling, and T. S. Spengler, "Upgrade auctions in build-to-order manufacturing with loss-averse customers," *European Journal of Operational Research*, vol. 250, no. 2, pp. 470–479, 2016.

[26] A. K. Chatterji, M. Findley, N. M. Jensen, S. Meier, and D. Nielson, "Field experiments in strategy research," *Strategic Management Journal*, vol. 37, no. 1, pp. 116–132, 2016.

[27] M. L. Markus, "Datification, organizational strategy, and IS research: what's the score?," *The Journal of Strategic Information Systems*, vol. 26, no. 3, pp. 233–241, 2017.

[28] N. Papadopoulos, L. Hamzaoui-Essoussi, and A. El Banna, "Nation branding for foreign direct investment: an Integrative review and directions for research and strategy," *The Journal of Product and Brand Management*, vol. 25, no. 7, pp. 615–628, 2016.

[29] T. Wrona and C. Sinzig, "Nonmarket strategy research: systematic literature review and future directions," *Journal of Business Economics*, vol. 88, no. 2, pp. 253–317, 2018.

[30] D. E. Kim, "Strategy research upon developing models for a community art and networking - with focus on the example of the Northern Area of Gyeonggi-do Province," *Journal of Digital Design*, vol. 17, no. 1, pp. 11–20, 2017.

[31] H. M. Li, G. J. Li, and H. F. Shi, "Current situation of civil awareness of undergraduate and promotion strategy research," *International Journal of Cognitive Research in Science Engineering and Education*, vol. 5, no. 1, pp. 1–6, 2017.

[32] E. Danneels, "Survey measures of first- and second-order competences," *Strategic Management Journal*, vol. 37, no. 10, pp. 2174–2188, 2016.

[33] S.-B. Tsai, R. Saito, Y.-C. Lin, Q. Chen, and J. Zhou, "Discussing measurement criteria and competitive strategies of green suppliers from a Green law Perspective," *Proceedings of the Institution of Mechanical Engineers-Part B: Journal of Engineering Manufacture*, vol. 229, no. S1, pp. 135–145, 2015.