International Journal of Innovative Computing, Information and Control
Volume **19**, Number **4**, August **2023**
ICIC International ©2023 ISSN 1349-4198
pp. 1221–1238

A COLLABORATIVE ALLOCATION APPROACH FOR SORTING RESOURCES IN RURAL COLD CHAIN WAREHOUSES

Di Wang, Haiyang Shi, Lijun Sun* and Yutian Du
School of Economics and Management
Dalian University of Technology
No. 2, Linggong Road, Ganjingzi District, Dalian 116023, P. R. China
{ wd; haiyang; duyt2022 }@mail.dlut.edu.cn; *Corresponding author: slj@dlut.edu.cn

Received October 2022; revised February 2023

Abstract. *In the warehousing procedure of fresh agricultural products in the rural cold chain warehouses, preliminary preservation is typically achieved through sorting, grading, and packaging in the sorting procedure. The allocation decision of sorting resources is crucial and complicated in the sorting procedure, as it not only reduces the loss of fresh agricultural products after harvesting but also facilitates operation cost savings in subsequent transportation procedures. Inappropriate resource allocation can result in excessive sorting time and a failure to process new sorting orders arriving randomly. Moreover, the excessive sorting time often leads to the decline of the quality of fresh agricultural products, which causes significant losses to both cold chain enterprises and rural farmers. This paper constructs a Multi-Agent Reinforcement Learning System for Multi-type Sorting Resources Collaborative Allocation (MARLS-MSRCA) to make the collaborative allocation of sorting resources in rural cold chain warehouses. In the MARLS-MSRCA, a collaborative interaction mechanism of sorting resource agents is proposed based on the analysis of the sorting procedure and the sorting resources involved. This mechanism obtains the strategy of collaborative allocation of multi-type sorting resources through the interactive learning between agents and the continuous interaction between agents and the sorting environment. Finally, the numerical experiments verified that the average order completion time and sorting cost in the rural cold chain warehouses can be greatly reduced by the proposed MARLS-MSRCA, which also demonstrates that the proposed approach can effectively promote the collaboration of the sorting procedure in rural cold chain warehouses.*

Keywords: Rural cold chain warehouses, Fresh agricultural products, Sorting and grading agents, Collaborative resource allocation, Multi-agent reinforcement learning

1. **Introduction.** Fresh agricultural products are perishable, vulnerable [1] and have high requirements for preservation [2]. "The report of the Planning, Design and Research Institute of the Ministry of Agriculture and Rural Affairs of the People's Republic of China in 2021" pointed out that the fresh agricultural products' loss rate after harvest in China is as high as 15%-25%, which leads to an annual loss of nearly 200 million tons. However, fresh agricultural products have to be harvested during a short period, and they cannot be sold immediately after harvesting due to the long distance between the rural areas and consumers. Therefore, after harvesting, these products need to be stored into the cold chain through the sorting procedure immediately in order to reach consumers far from the rural area through cold chain transportation. The sorting procedure is a key step in reducing the loss of the fresh agricultural products' quality. Moreover, if the sorting accuracy of fresh agricultural products is more than 90%, it can save 35% of the packaging and cold storage costs for subsequent procedures [3]. Therefore, timely and

DOI: 10.24507/ijicic.19.04.1221

efficient sorting of fresh agricultural products cannot only reduce effectively the loss of quality after harvesting but also save the cost of rural cold chain warehouses greatly.

However, the sorting procedure of fresh agricultural products still does not receive sufficient attention in today's rural cold chain warehouses. For example, fresh agricultural products are mostly sorted at room temperature for a long time, which results in an exponential decline in quality with exposure time [4]. Especially during the fruit ripening season, the product loss caused by unreasonable sorting operations will be more serious. There are two main reasons for the great loss. On the one hand, the data of the third agricultural census shows that the number of farmers accounts for more than 98% of the main agricultural operators in China. The harvested fresh agricultural products are diverse and scattered, resulting in the diverse and scattered demands for sorting fresh agricultural products. On the other hand, the unit price of fresh agricultural products is relatively low, while the resources required for sorting are relatively expensive. Enterprises have very limited investment in sorting procedure, and can only afford to purchase a few sorting machines and hire a small number of packing workers. Furthermore, the sorting resources are usually allocated once a day based on manual experience. It is difficult for a few sorting resources to fulfill the randomly arrived sorting demands and process the agricultural products into the cold warehouse timely when they arrive. In summary, in the sorting procedure of the rural cold chain warehouses, with the dynamic arrival of sorting demands of fresh agricultural products, the limited resources in the sorting warehouse must be collaboratively allocated to avoid excessive retention time in the sorting procedure.

The contributions of our research can be summarized as follows. Firstly, we establish an MARLS-MSRCA framework for optimizing the collaborative allocation of the multi-type sorting resources in rural cold chain warehouses. Secondly, we design a collaborative interaction mechanism for sorting resource agents, which aims to drive the interactive learning between agents and the environment in MARLS-MSRCA. Thirdly, based on the proposed collaborative interaction mechanism, a multi-agent reinforcement learning collaborative decision-making approach is developed to obtain the final collaborative allocation strategy. Lastly, the numerical experiments based on practical cases demonstrate that the proposed approach can effectively promote the collaboration of sorting procedure in rural cold chain warehouses.

The rest of this paper is organized as follows. Section 2 reviews the related literature. In Section 3, the problem of collaborative allocation of sorting resources in rural cold chain warehouses is described and related notations are given. Section 4 constructs the MARLS-MSRCA and proposes a collaborative decision-making approach based on the Multi-Agent Deep Deterministic Policy Gradient (MADDPG) framework. Section 5 presents the numerical experiments and results. The conclusions and future work are stated in Section 6.

2. **Literature Review.** In recent years, scholars have directed their attention towards optimizing the sorting procedure to address the problems associated with quality deterioration in the supply cold chain of fresh agricultural products. Zhu and Li [5] proposed an intelligent sorting method that utilizes machine vision technology to analyze multiple features, achieving in a sorting rate of over 95% for spherical fruits and vegetables. Yu et al. [6] designed the target positioning function and sorting strategy of the fruit sorting robot based on image processing technology and proved the effectiveness of the sorting strategy. Wang et al. [7] developed an automatic sorting system for fresh white button mushrooms based on image processing technology, which realized the non-destructive sorting of fresh white button mushrooms. Zhang et al. [3] developed a sorting system for apples

and verified that the grading accuracy of the system can reach more than 99%, which is enough to support the real-world commercial operation. Leung et al. [8] introduced a comprehensive analysis method for grading fresh agricultural products and established a standard operating framework for food sorting, which improved the efficiency of fruit sorting in practice. At present, the research on sorting of fresh agricultural products mainly focuses on the optimization of sorting system technologies such as visual recognition and internal non-destructive testing using machine learning and deep learning. The primary objective is to enhance operation accuracy of the sorting system, which provides technical support for the machine sorting of various fresh products. However, most of the research only briefly designs the process of the sorting procedure, without consideration of the collaborative allocation of multi-type sorting resources. In reality, the sorting system for fresh agricultural products consists of many links and involves a variety of resources. If we solely concentrate on improving the sorting accuracy and allocate a fixed amount of resources for each order without consideration of the specific situation, these resources may not be fully utilized in most situations, resulting in congestion between links. This indicates that the sorting system efficiency for fresh agricultural products in real-world settings will still be suboptimal without considering resource allocation for sorting. Consequently, it is necessary to investigate the collaborative allocation of sorting resources in the sorting system further. However, it is difficult to establish a system for fresh agricultural products to allocate sorting resources collaboratively with the sorting orders of random arrival in the practical application.

Existing studies mainly focus on the collaborative resource allocation of sorting systems in warehouses that are not specifically designed for the rural cold chain. These studies have investigated the sorting resources such as shelves, AGVs, and pickers. Lambrechts et al. [9] studied the human-machine collaboration problem, considering human factors in the warehouse sorting system. Yu et al. [10] studied the problem of multi-AGV task assignment and vehicle routing optimization in the sorting bin. Jiang and Huang [11] optimized the Robotic Mobile Fulfillment System (RMFS) with consideration of the order delivery requirements. Fager et al. [12] established a sorting system with human-machine collaborative optimization for minimizing the cost objective related to workers, equipment, and quality. Zhou et al. [13] proposed an information-based multi-criteria indexing method to optimize the business intelligence system of semi-automated sorting facilities for the problem of determining the order of cargo loading and unloading. Zeng et al. [14] designed a set of data representation methods for industrial sorting operations by combining Deep Reinforcement Learning (DRL) and Genetic Algorithm (GA). These studies provide possible solutions to establish a collaborative resource allocation system for sorting resources. However, since the quality of fresh products deteriorates exponentially over time, the internal process has a higher requirement for timeliness. This differs from a typical warehouse sorting system and requires further consideration of how to address the high timeliness requirements. Additionally, the external demands in the sorting system of rural cold chain warehouses are stochastic and make the collaborative allocation of sorting resources more complicated. Therefore, in this paper, we need to design a new collaborative resource allocation system that can adapt to the multi-type sorting resources of rural cold chain warehouses.

In summary, the new collaborative resource allocation system that we expect to design should be able to allocate multi-type sorting resources collaboratively in the complex situations of randomly arriving orders and orders with high timeliness requirements. And the Multi-Agent System (MAS) is composed of multiple agents, which can adapt to the complex external environment better and accomplish the system tasks than a single agent, so it is suitable for the collaborative allocation of system resources in a random dynamic environment. Tong et al. [15] used multi-agent system to reduce the impact of dynamic uncertainty on the collaborative allocation of the original surgical resources. Aiming at the problems of difficult collaborative control and low efficiency of agricultural agent groups, Gong et al. [16] studied the task allocation of heterogeneous agricultural agent groups based on a dynamic stimulus-response model. With the development of reinforcement learning algorithms, some researchers have applied reinforcement learning algorithms to multi-agent system and formed Multi-Agent Reinforcement Learning System (MARLS). MARLS enables agents to accomplish more complex tasks through interaction and collaborative decision-making in higher-dimensional and dynamic real-world scenarios. According to research in recent years, MARLS is gradually applied in the field of resource allocation [17]. Nguyen et al. [18] used non-cooperative real-time MARLS to solve the energy-saving power co-allocation problem while satisfying the QoS constraints in D2D communication. Zhang et al. [19] carried out online optimization of a household energy management system based on asynchronous deep reinforcement learning, which can provide real-time feedback for power users. Ying et al. [20] proposed an MADRL-based adaptive control system for subway collaborative operation with a flexible train combination, which has advantages in the flexible arrival of trains. However, the application of MARLS in the field of resource allocation is still in the preliminary phase and most of the existing MARLSs are designed with a single type of agent. However, the sorting resources in rural cold chain warehouses are multi-type resources and located in the sorting line with a fixed sequence. In the existing MARLS, it is not possible to guarantee the convergence of multi-type agents' collaborative decision-making. Therefore, it is necessary to develop a new multi-agent reinforcement learning system for multi-type resources to realize the collaborative allocation of sorting resources in rural cold chain warehouses.

3. Problem Statement and Notations.

3.1. **Problem statement.** In order to ensure that fresh agricultural products can be sorted and processed as soon as possible after harvesting, the rural cold chain warehouses are usually located near the product harvesting sites with convenient transportation, which can provide sorting services for scattered farmers in surrounding rural areas. Sorting services for fresh agricultural products generally include the following procedures. Firstly, the fresh agricultural products that need to be sorted and processed arrive at the sorting warehouse. The sorting warehouse will establish a corresponding order including the information of the farmers to which the fresh agricultural products belong, the weight of products that need to be sorted, the grading standards of fresh agricultural products, and the grading distribution. Secondly, fresh agricultural products are cleaned for preliminary preservation, waiting for sorting. Thirdly, the fresh agricultural products are sorted and graded through the sorting procedure. During the sorting procedure, rotten products are screened out while other products are graded and packaged according to the grading standards. Then the packaged products will be stored in the refrigerated warehouse for preservation. The sorting resources involved in the cold chain warehouses in rural areas are shown in Figure 1, mainly including machine resources and human resources. The machine resources include sorting machines, grading machines, and packaging machines, and the human resources mainly refer to the packaging workers.

There are several sorting lines in one sorting warehouse. Each sorting line includes several sorting machines, grading machines, packaging machines, and some packaging workers. The number of channels of each machine resource has a fixed upper limit and the channels have a fixed processing speed. Therefore, the processing efficiency of the machine can be adjusted by switching the channels. Due to the different processing speeds

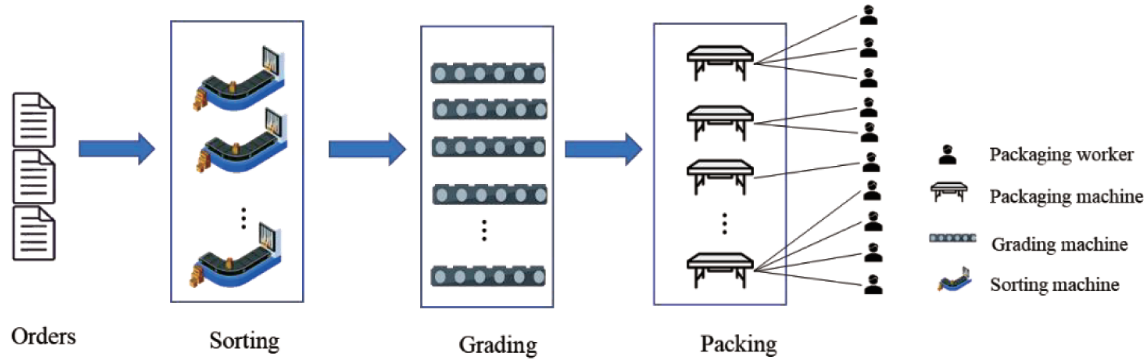

FIGURE 1. Sorting procedure flow chart

of different machine channels, in order to be able to synergize the processing efficiency between the links as much as possible, the channels of the three kinds of machines have the following constraints: the maximum number of sorting machine is less than the maximum number of packing machine, and the maximum number of packing machine is less than the maximum number of grading machine. Although the upper limit of sorting resources is known, we cannot always turn on all the channels of each machine since the amount and grade distribution of each order are different. Therefore, it is necessary to make specific decisions on the allocation of various resources according to the randomly arriving sorting orders. Specifically, it is necessary to give an appropriate strategy for the collaborative allocation of sorting resources, including the following decisions: the number of channels opened by the sorting machines related to the total amount of products to be sorted, the number of channels opened by the grading machines according to the grade distribution of orders, the number of packaging machines corresponding to each grading machine, and the workers allocated to each packaging machine. This is a multiple-resources collaborative allocation decision problem under the random arrival of sorting orders, and the decision objective is to minimize the total sorting cost of orders with limited sorting resources. Based on the setup of the actual sorting system, we have the following settings for our problem.

1) Sorting orders arrive randomly and are sequentially processed.

2) Due to the same delivery deadline and type of agricultural products, an individual order cannot be split.

3) The sorting weight and the grade distribution of each order are known.

4) If the sorting time exceeds the maximum completion time of the order, the order will lose its original economic value and will be recorded as an unfulfilled order.

5) Channels with the same type of machine resource have the same speed, and the processing efficiency of the procedure is affected by the number of processing channels opened.

6) The packaging speed is proportional to the number of packaging workers.

7) The buffer areas before and after the machines have unlimited capacity.

3.2. **Notations.** Relevant mathematical symbols are shown in Table 1.

The goal of collaborative allocation for sorting resources of cold chain warehouses in rural areas is to minimize the total sorting cost of orders with limited resources, which includes machine operation cost, labor cost, and order tardiness cost as follows:

$$Min\, c_j = \begin{cases} c1_j + c2_j, & td_j \leq \max td_j \\ c1_j + c2_j + p_j, & td_j > \max td_j \end{cases} \quad j \in J,\ i \in I \tag{1}$$

$c1_j$ is the operation cost of all machines for order j, which is the sum of the processing time cost of machine i, the fixed cost proportional to the number of channels allocated by machine i, and the operation cost per channel of machine i. $c2_j$ is the labor cost for order j, which is determined by the number of packing workers allocated and the unit labor cost of one packing worker. So $c1_j$ and $c2_j$ are calculated as

$$c1_j = \sum_1^I tm_{ij} \times k_i \times b_{ij}, \quad j \in J, i \in I \tag{2}$$

$$c2_j = km \times m_j, \quad j \in J \tag{3}$$

The completion time td_j of order j begins when order j enters the sorting warehouses and ends when the packaging is completed, including the staging time in front of the machine, the processing time of the machine, and the manual packaging time. It can be calculated by Equation (4).

$$td_j = \sum_1^n tw_{ij} + \sum_1^n tm_{ij} + tp_j, \quad j \in J, i \in I \tag{4}$$

The deterioration of fresh agricultural products primarily depends on the sorting time and temperature. Let ρ be the rate of decrease in the quality of fresh agricultural products over time, and the expression formula of the quality of raw agricultural products in time

Table 1. Symbols and variables

Symbol	Variables
J	Set of order j
I	Set of machine i
u	Classified number of machine channels
B_i	Maximum number of channels allocated by machine i
M	Total number of packaging workers
MP	Upper limit of the number of workers per packing machine
k_i	Unit operation cost per channel of machine i
km	Unit labor cost per packing worker
f_{ij}	Idle condition of machine i when order j arrives
$c1_j$	Machine operation cost of order j
$c2_j$	Labor cost of order j
c_j	Total sorting cost of order j
w_j	Sorting weight of order j
p_j	Value of order j
$\max td_j$	Maximum completion time of order j
td_j	Completion time of order j
tw_{ij}	Staging time of order j in machine i
tm_{ij}	Processing time of order j in machine i
tp_j	Manual packing time of order j
ttd_{ij}	Remaining processing time from the maximum completion time when order j arrives at machine i
b_{ij}	Number of channels allocated by machine i for order j
b_{uij}	Number of channels allocated by machine i for order j according to classification u
m_j	Number of packing workers allocated for order j
m_{uj}	Number of packing workers allocated for order j according to classification u

t is $M_t = M_S \cdot e^{\rho t}$ [4]. Therefore, the processing of fresh agricultural products must be completed before the minimum quality M_l is compromised. If the processing time of order j exceeds the maximum processing time $\max td_j$, then the product quality is lower than the required minimum quality M_j, which means the product cannot be sold and the order j will lose its original value p_j. The value p_j can be calculated by Equation (5).

$$p_j = \begin{cases} p_j & td_j \le \max td_j \\ 0 & td_j > \max td_j \end{cases}, \ j \in J \tag{5}$$

On a sorting line, the number of sorting machine channels, the packing machine channels, and the grading machine channels is gradually increased as shown in Equation (6):

$$B_1 < B_3 < B_2 \tag{6}$$

The number of each resource allocated is an integer value and cannot exceed the maximum number of that resource. The channels of the grading machine and the packing machine that will be turned on can be determined according to the grade distribution. Furthermore, the number of resources allocated to each channel must also be an integer.

$$b_{2j} = \sum_{1}^{u} b_{u2j}, \ b_{3j} = \sum_{1}^{u} b_{u3j} \tag{7}$$

$$b_{1j} \le B_1, \ b_{2j} \le B_2, \ b_{3j} \le B_3 \tag{8}$$

$$m_j \le M, \ M_{uj} \le MP \tag{9}$$

$$b_{1j}, b_{u2j}, b_{u3j}, m_{uj}, B_i, M, MP \in N \tag{10}$$

4. **Multi-Agent Reinforcement Learning System for Multi-Type Sorting Resources Collaborative Allocation.** In the rural cold chain warehouses, the arrival of sorting orders has strong randomness as the farmers are scattered in rural areas. In the meanwhile, the sorting procedures of these orders have narrow time windows due to the rapid deterioration of fresh agricultural products. To address the collaborative allocation problem of multi-type sorting resources in rural cold chain warehouses, the collaborative allocation decision among multi-type resources needs to be generated with consideration of the time-sensitive and randomly arriving orders. As mentioned above, the MARLS can accomplish more complex decision-making tasks through interactions and decisions between multi-type agents and the environment. Therefore, this paper designs an MARLS-based system and proposed an MADDPG-based algorithm for the collaborative allocation problem of multi-type sorting resources in rural cold chain warehouses. This section is organized as follows. Firstly, we established the MARLS-MSRCA and defined the state space, action space, and reward function of each agent. Secondly, we constructed a collaborative interaction mechanism to enable interaction among sorting resource agents in the MARLS-MSRCA. Finally, we designed a multi-agent collaborative decision algorithm based on the framework of MADDPG algorithm for MARLS-MSRCA. Using the collaborative decision algorithm, the MARLS-MSRCA can generate a set of collaborative allocation decisions for multi-type sorting resources in a dynamic environment with random arrival of orders and timely sorting procedure requirements.

4.1. **Multi-agent reinforcement learning for multi-type sorting resources collaborative allocation.** In MARLS, agents make joint decision action $A = \{a_1, a_2, \dots, a_n\}$ in the joint state S_t, and communicate with the sorting environment and each other to obtain a joint reward $R = \{r_1, r_2, \dots, r_n\}$, and then transfer to the state S_{t+1}. In the MARLS-MSRCA designed in this paper, multi-type agents generate the collaborative allocation strategy through continuous interactions with the environment and other agents, eventually realizing the decision optimization of the multi-agent system [21]. The MARLS-MSRCA takes the sorting resources that need to be allocated collaboratively in the rural cold chain warehouses as agents, which include sorting machines, grading machines, packaging machines, and packaging workers. However, since the processing time of the packaging machine depends on the packaging time of the workers, we take the packaging machine and all the packaging workers working on it as one packaging agent. Therefore, as shown in Figure 2, the three types of agents involved in the MARLS-MSRCA include sorting agent A_1, grading agent A_2, and packaging agent A_3.

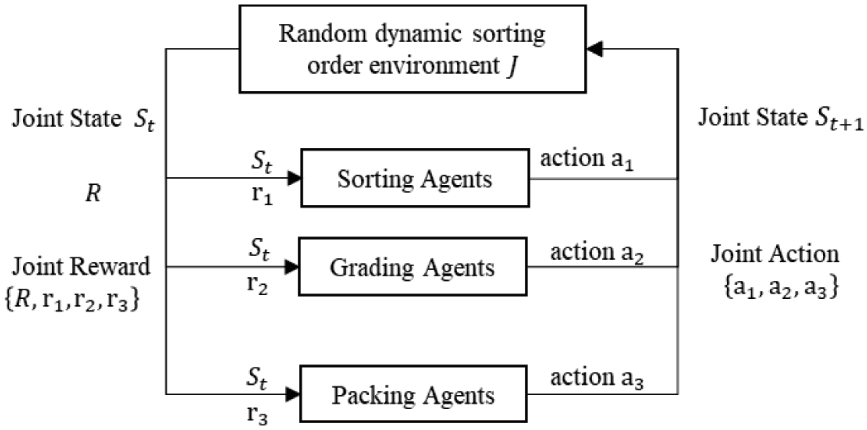

FIGURE 2. Multi-agent reinforcement learning system for multi-type sorting resources collaborative allocation

4.1.1. *State space.* The sorting environment state S can be defined as $S = \{J, o_1, o_2, o_3\}$ [22], which is composed of the orders' state J and the state of three types of agents $\{o_i, i = 1, 2, 3\}$.

The order state space J includes the orders' index j, the maximum completion time $\max td_j$ of order j, the sorting weight w_j of order j, and the value p_j of order j. They can be specifically represented as follows:

$$J = \{j, \max td_j, w_j, p_j\} \tag{11}$$

The state space of the three types of agents is specifically expressed as follows.

A_1 represents the sorting agent. The state of A_1 includes the staging time tw_{1j} and the processing time of sorting machine tm_{1j}, the remaining processing time ttd_{1j} calculated as the difference between the maximum completion time and the time when order j arrives at the sorting machine, the maximum number of channels allocated by sorting machine B_1 and the idle condition of sorting machine $f_{1j} = \begin{cases} 0 & \text{not free} \\ 1 & \text{free} \end{cases}$. The observable state of the sorting agents can be expressed as

$$o_1 = \{tw_{1j}, tm_{1j}, ttd_{1j}, B_1, f_{1j}\} \tag{12}$$

A_2 represents the grading agent. The state of A_2 includes the staging time tw_{2j} and the processing time of grading machine tm_{2j}, the remaining processing time ttd_{2j} calculated as the difference between the maximum completion time and the time when order j arrives at grading machine, the maximum number of channels allocated by the grading machine B_2 and the idle condition of grading machine $f_{2j} = \begin{cases} 0 & \text{not free} \\ 1 & \text{free} \end{cases}$. The observable state of the grading agents can be expressed as

$$o_2 = \{tw_{2j}, tm_{2j}, ttd_{2j}, B_2, f_{2j}\} \tag{13}$$

A_3 represents the packing agent. The state of A_3 includes the staging time tw_{3j} and the processing time of packing machine tm_{3j}, the remaining processing time ttd_{3j} calculated as the difference between the maximum completion time and the time when order j arrives at packing machine, the total number of packaging workers M, the maximum number of channels allocated by packing machine B_3 and the idle condition of packing machine $f_{3j} = \begin{cases} 0 & \text{not free} \\ 1 & \text{free} \end{cases}$. The observable state of the packing agents can be expressed as

$$o_3 = \{tw_{3j}, tm_{3j}, ttd_{3j}, M, B_3, f_{3j}\} \tag{14}$$

4.1.2. *Action space.* All resource agents can observe the sorting environment state S in the warehouse and then make joint decision-making actions. The set of joint actions of resource agents is denoted as $a = (a_1, a_2, \dots a_n)$, which are explained as follows.

1) Sorting agent A_1. A_1 decides the number of channels of sorting machine b_{1j} based on the sorting environment state S.

$$a_1 = \{b_{1j}\} \tag{15}$$

2) Grading agent A_2. According to the classification u from the grading distribution of order j, the channels are divided into u classes. A_2 needs to decide the number of classified channels of grading machine b_{u2j} based on the sorting environment state S.

$$a_2 = \{b_{u2j}\} \tag{16}$$

3) Packing agent A_3. According to the classification u from the grading distribution of order j, the channels are divided into u classes. Then, based on the sorting environment state S, A_3 decides the number of classified channels of packaging machine b_{u3j} and the corresponding number of packaging workers m_{uj}.

$$a_3 = \{b_{u3j}, m_{uj}\} \tag{17}$$

4.1.3. *Reward function.* The transition of the environment state is determined by the previous state of the environment and the joint actions of multi-type agents. Each agent's action is affected by the reward function denoted as $r_i = r_i(s_i, a_i, s_i')$. However, since the reward function of each agent can only be observed by the agent itself, it is necessary to design a joint reward function R and a single reward function r_i separately to guide the agent's action.

The negative value of the order completion cost in the sorting procedure is set as the joint reward of multi-agents as shown in Equation (18), which means that the lower total sorting cost of order j leads to larger joint multi-agent rewards. The agent aims to maximize the reward through parameter updating and provides guidance for the update of the subsequent reinforcement learning model.

$$R = -c_j = \begin{cases} -(c1_j + c2_j) & td_j \le \max td_j \\ -(c1_j + c2_j + p_j) & td_j > \max td_j \end{cases} \tag{18}$$

Moreover, each agent has its own reward, which is to minimize the operation cost in their respective procedures. Therefore, the negative number of the operation cost of each agent is used as the reward in Equations (19) to (21).

Reward function r_1 of A_1:

$$r_1 = -k_1 \times td_{1j} \times b_{1j} \tag{19}$$

Reward function r_2 of A_2:

$$r_2 = -k_2 \times td_{2j} \times b_{2j} \tag{20}$$

Reward function r_3 of A_3:

$$r_3 = -td_{3j} \times (k_3 \times b_{3j} + km \times m_j) \tag{21}$$

4.2. **Multi-agent reinforcement learning collaborative decision-making approach based on MADDPG framework.** MADDPG algorithm [23] is an MARLS algorithm based on Deep Deterministic Policy Gradient (DDPG) algorithm [24]. The DDPG algorithm combines the Actor-Critic algorithm [25] with the Deep Q-Network (DQN) algorithm [26] and performs well in action decisions. While inheriting the advantages of DDPG, MADDPG algorithm proposes a framework of centralized training and distributed application to solve the environmental instability problem of MARLS. In order to simplify the policy training process, the MADDPG algorithm allows each agent to learn from global information in the centralized training process. In the application process, the MADDPG algorithm allows each agent to decide based on local information and make the environment of MARLS more stable by maximizing the set objective. The MADDPG algorithm not only solves the problem of a non-stationary environment but also allows the reward function of each agent to be different, which can enhance the effectiveness of collaborative decision-making in the multi-agent system. Furthermore, MADDPG is a state-of-the-art algorithm to solve the MARLS and has considerable performance over other traditional algorithms, especially the problems of resource allocation [27]. Therefore, based on the MADDPG algorithm framework, this paper proposes a multi-agent reinforcement learning collaborative decision-making approach to generate the action decision set of the MARLS-MSRCA.

4.2.1. *Design of collaborative interaction mechanism for sorting resource agents.* In the MADDPG framework, the multi-agent estimates the action decision of other agents according to the global information and makes its own decision considering the estimated strategy. However, in the case of disordered interaction of multi-type resource agents, it is easy to cause non-convergence of collaborative decision-making. We found that in the decision-making process of the multi-type sorting resource agents, the action decision has the order of "sorting-grading-packing". The action decision and the state after the decision of the agent will affect the action decision of the subsequent agent. For example, a grading agent's action decision can guide its subsequent packaging agent's action decision. Therefore, this paper sets the decision-making sequence of the sorting resource agents as $A_1 \to A_2 \to A_3$, and designs a centralized training coordination mechanism and shared parameter network between agents. This meets the requirements of resource allocation rules in real decision-making, and at the same time, it can improve the communication efficiency between the sorting resource agents, reduce the duration of centralized training and solve the problem that the collaborative decision-making of multi-type agents is difficult to converge.

Through the centralized training process, we design a parameter-sharing network to facilitate the interaction between the sorting resource agents. The parameter-sharing network mainly includes the observable state $o_{all} = \{o_1, o_2, o_3\}$ of agents and the order state J of the order is processed. The description of the collaborative interaction mechanism of sorting resource agents through the parameter network is shown in Figure 3. Sorting orders enter into the MARLS-MSRCA according to the first-in, first-out sorting principle. The status feature J of the order to be sorted is input into the MARLS-MSRCA as part of the environment and stored in the shared parameter network. Then the MARLS-MSRCA starts the collaborative allocation decision of sorting resources for the order. First, the sorting agent A_1 makes a decision action a_1 based on all current observable states o_{all} and order status J. The next state of the sorting agent can be obtained from o_1'. Since

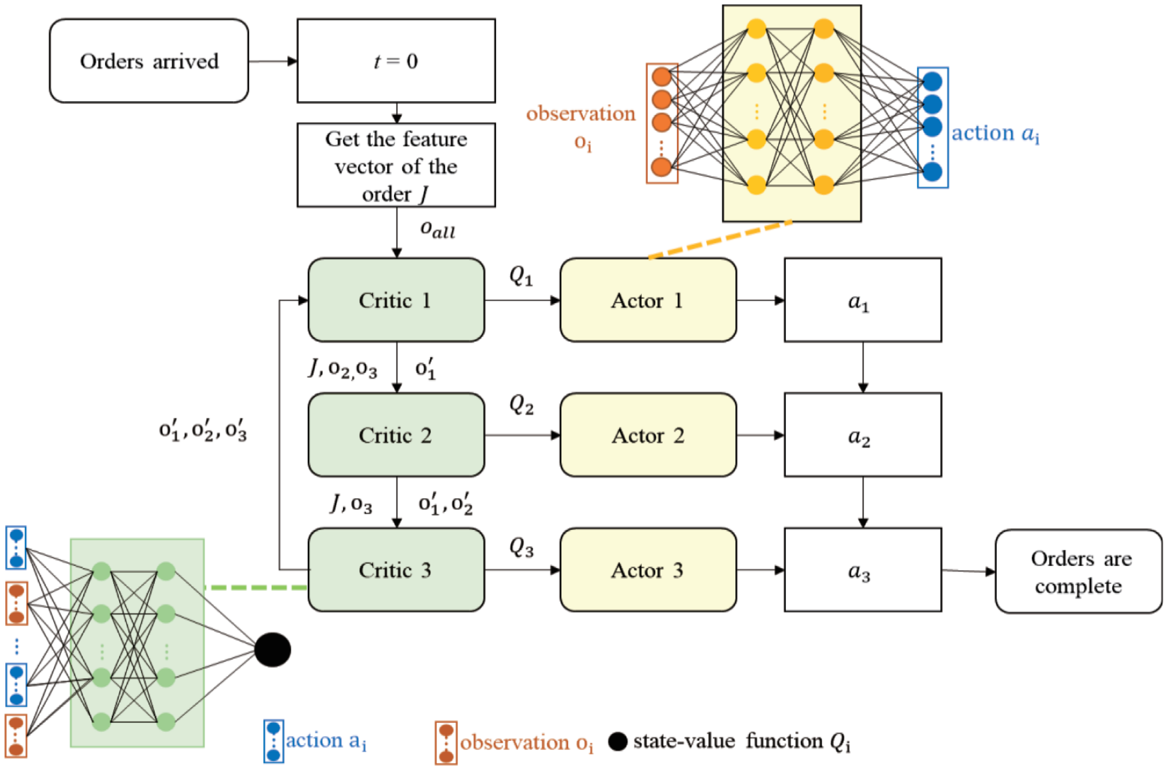

FIGURE 3. (color online) The collaborative interaction mechanism of sorting resource agents

other agents have not yet made a decision, the environment will not actually perform the actions of A_1 at this time, but update o_1' to the parameter-sharing network to provide a new decision state for the other agents to make subsequent decisions. Then, the grading agent A_2 will make a decision a_2 based on the new observable state $o_{all} = \{o_1', o_2, o_3\}$, and the order feature J. Similarly, the state of the packaging agent A_3 when making a decision a_3 is based on the observable state $o_{all} = \{o_1', o_2', o_3\}$ and order feature J. Finally, execute the sorting joint action $a = \{a_1, a_2, a_3\}$, and output the sorting resource collaborative configuration scheme of order J. At the same time, the observable state of the shared parameter network is updated to $o_{all} = \{o_1', o_2', o_3'\}$. So far, the interactions between agents are completed.

4.2.2. *Training and updating.* Each agent has a corresponding policy network named Actor-network, a value network named Critic-network, and their target networks, respectively. In Actor-network, input a certain o_i and output a certain action $a_i = \mu(o_i, \theta_i)$. In Critic-network, the input is the global state $S = \{J, o_1, \ldots, o_n\}$ and the actions of all agents $a = \{a_1, \ldots, a_n\}$. The output is the "state-value" function $Q_i = \{s, a, \omega_i\}$. α is the learning rate. α_ω and α_θ represent the learning rates of the two networks, respectively. r is the reward corresponding to the action a_t. The target network of the Critic-network gives $Q(s_{t+1}, a_{t+1})$. $\nabla_\omega Q(s_t, a_t)$ is the update gradient of the Critic-network and $\nabla_\theta \log \pi_\theta(a_t|s_t)$ is the update gradient of the Actor-network. To stabilize the learning process, both the Actor-network and the Critic-network set their own target network. The updated formulas are as follows:

$$Q^\mu(s_t, a_t) = R(s_t, a_t) + \gamma Q^\mu(s_{t+1}, a_{t+1}) \tag{22}$$

$$\delta = r_t + \gamma Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t) \tag{23}$$

$$\omega_{t+1} = \omega_t + \alpha_\omega \delta_t \nabla_\omega Q(s_t, a_t) \tag{24}$$

$$\theta_{t+1} = \theta_t + \alpha_\theta \nabla_\theta \log \pi_\theta(a_t|s_t)\delta_t \tag{25}$$

4.2.3. *Multi-agent reinforcement learning collaborative decision-making algorithm.* The process of the multi-agent collaborative decision-making algorithm is described as follows. Firstly, the algorithm requires the input of the random dynamic sorting state S, including the order state J and the initial state O of the resource agents. Secondly, driven by the incoming orders, the resource agents will store the interactive data and continually generate collaborative allocation strategies for multi-type sorting resources during centralized training. By receiving the strategies of other resource agents continuously, the resource agents can finally estimate the strategies of other agents after centralized training. Then, after the centralized training, each resource agent can quickly output its own collaborative allocation decision a_i for multi-type sorting resources when making distributed decisions. Finally, after inputting the sorting state S, the resource agents can generate the corresponding decision program quickly. The pseudo-code of the algorithm is shown in Table 2.

Table 2. Multi-agent reinforcement learning collaborative decision-making algorithm

Input: Order state J, Multi-type sorting resource agent state O
Output: Joint decision-making of multi-agents A
for episode = 1 to M **do**
Initialize a random process N;
Initial state $x = (J, O)$;
  for $t = 1$ to max episode length **do**
    for agent i, select action $a_i = \mu_{\theta_i}(J, o_i)$
      Execute actions $a = (a_1, a_2, \ldots, a_n)$ and observe reward and new state x'
      Store (x, a, r, x') in replay buffer D
      $x \leftarrow x'$
      for agent $i = 1$ to n **do**
        Sample a random minibatch of S samples (x_d, a_d, r_d, x'_d) from D
        Set $y_d = r_{id} + \gamma Q_{\mu' i}\left(x'_d, a'_1, a'_2, \ldots, a'_n\right)\big|_{a'_k = \mu'_k(o_{kd})}$
        Update Critic by minimizing the loss $L(\theta_i) = \frac{1}{S}\sum_d (y_d - Q_{\mu i}(x_d, a_{1d}, \ldots, a_{nd}))^2$
        Update Actor using the sampled policy gradient:
$$\nabla_{\theta_i} D \approx \frac{1}{S}\sum_d \nabla_{\theta_i}\mu_i(o_{id})\nabla_{a_i} Q_{\mu i}\left(x_j, a_{1d}, \ldots, a_{nd}\right)\Big|_{a_i}$$
      end for
        Update target network parameters for each agent i:
$$\theta'_i \leftarrow \tau\theta_i + (1-\tau)\theta'_i$$
  end for
end for

5. Numerical Experiment.

5.1. **Simulation data.** To verify the validity of the MARLS-MSRCA, we simulate the actual rural cold chain warehouse environment and a random arrival order scenario. Based on the actual sorting procedure of an apple cold chain sorting warehouse in Shaanxi Province, China, the experiment is designed to simulate the real scenario of the sorting system in rural cold chain warehouses. The certified output of the apple production cooperative to which the sorting warehouse belongs is 800 t. That means apple production cooperative can provide sharing sorting services to others. Therefore, there are two main types of sorting orders: one type comes from surrounding scattered farmers, with a harvest volume between 8.4 t to 16.8 t each time, while the other type comes from the rural cooperative to which the sorting warehouse belongs, with a harvest volume between 16 t to 32 t each time. The total harvest volume of apples in each batch of sorting orders is between 8.4 t to 32 t, and the average price is 3400 ¥/t [28].

In the simulation experiment, it is assumed the arrival of sorting orders follows an exponential distribution, and the weight of each order is randomly generated between 8.4 t to 32 t. The maximum completion time of an order is calculated as the difference between the quality requirement time and the randomly generated harvest time. Due to the seasonal fluctuations in the harvest of fresh agricultural products, two kinds of cases are tested: one with intensive arrival of orders and the other with sparse arrival of orders. It is worth noting that there is only one sorting line in the sorting warehouse. During the grading process, apples are divided into five sizes: 70, 75, 80, 90, and rotten apples. The percentage of rotten apples is kept below 10%. The distribution of grading is subject to Poisson distribution. The relevant indicators for sorting resources are listed in Table 3.

Table 3. Relevant indicators for sorting resources

Indicators	Value
The speed of the sorting machine channels	5 t/h
The speed of the grading machine channels	6 t/h
The speed of per packing worker	0.135 t/h
k_i	2 ¥/h
km	22.5 ¥/h
MP	7
B_1	3
B_2	15
B_3	10

Based on the above settings, this paper constructed a sorting simulation system of rural cold chain warehouses and proves the validity of MARLS-MSRCA through economic benefit analysis and convergence analysis. Firstly, we use the Daily Fixed Resource Allocation Strategy (DFRAS) in the real world as the benchmark of the economic benefit analysis. DFRAS is a manual strategy where operators make daily fixed resource allocations based on cumulative data from recent days, due to the inability to allocate sorting resources in real time based on the human experience. And we compare the relevant indicators of MARLS-MSRCA and the DFRAS to prove the economic benefit of MARLS-MSRCA. Secondly, we use the original MADDPG as the benchmark of convergence analysis and conduct a convergence analysis comparison between the original MADDPG and the MADDPG with our proposed interactive mechanism, which can demonstrate the superiority of our algorithm in terms of convergence. In the experimental system, the specific parameter settings are shown in Table 4 and the training case of each episode consists of random order data with fixed seeds.

5.2. **Validity analysis.**

5.2.1. *Economic benefit analysis.*
Through the economic benefit indicators obtained by numerical experiments, the results of MARLS-MSRCA and the DFRAS in the real world

TABLE 4. Experiment parameters setting

Parameter	Value
Number of hidden layers	4
Number of hidden layer neurons	64
Hidden layer activation function	Relu
Output layer activation function	Softmax
Optimizer	Adam
γ	0.95
Span	20000
ϵ	0.95
α_ω	0.01
α_θ	0.05
Replay memory size	512
Minibatch size	128
Due date tightness-high	$U[1,2]$
Due date tightness-low	$U[1,3]$

are compared and analyzed. The analytical indicators include the order completion rate, the average order sorting time, and the average order sorting cost. The order completion rate is defined as the ratio of the number of completed orders to the number of randomly arrived orders under the condition of limited resources. The average order completion time is calculated as the total order completion time divided by the number of orders. At the same time, the average order sorting cost is calculated by dividing the total order operation cost by the number of completed orders. The total sorting cost of orders includes the machine operation cost and the labor cost. In the same order environment, it can be proved that MARLS-MSRCA can better deal with the randomly arrived orders by comparing two indicators: the average order completion time and order completion rate of MARLS-MSRCA and DFRAS. Moreover, it can be verified that the MARLS-MSRCA can save the warehouse operation cost by comparing the average order sorting cost of MARLS-MSRCA and DFRAS.

In the case of generating the same random orders, by comparing the average order sorting cost it can be concluded from Table 5 that the proposed resource collaborative allocation strategy of MARLS-MSRCA can reduce the operation cost of the sorting warehouse. The average order sorting cost is reduced by 14.38% in the case of intensive arrival of orders and 7.54% in the case of sparse arrival of orders. In particular, the high cost of packaging workers has been significantly reduced. At the same time, the order completion rate and average completion time of the MARLS-MSRCA are better than the benchmark, whether orders arrive intensively or not. Through the comparison of the above indicators, it can be proved that the MARLS-MSRCA has a great decision-making advantage in the face of randomly arriving sorting orders.

5.2.2. *Convergence analysis.* Convergence is an important technical indicator of the algorithms for MARLS. Therefore, a comparison of the convergence speed and the final convergence value is conducted between the original MADDPG algorithm and the MADDPG algorithm with the interaction mechanism during the training process, which can demonstrate the superiority of the MADDPG algorithm with the interaction mechanism in convergence with the same parameters setting. To facilitate visual comparison, the comparison between these two algorithms is shown in Figure 4, in which the horizontal axis reports the number of training episodes and the vertical axis reports the joint reward

TABLE 5. Related indicators for validity analysis

Indicators		Intensive order					Spare order				
		SM	GM	PM	PW	Sum	SM	GM	PM	PW	Sum
Average order sorting cost (¥)	MARLS-MSRCA	4.85	13.65	13.34	1308.56	1340.53	4.05	8.80	9.92	925.11	974.89
	DFRAS	6.14	11.63	20.15	1527.78	1565.70	4.09	7.73	13.47	1020.20	1045.49
Order completion ratio (%)	MARLS-MSRCA			84.98					91.43		
	DFRAS			82.35					85.72		
Average order completion time (h)	MARLS-MSRCA			3.667					2.363		
	DFRAS			3.875					2.591		

Note: SM means sorting machine, GM means grading machine, PM means packing machine, and PW means packing workers.

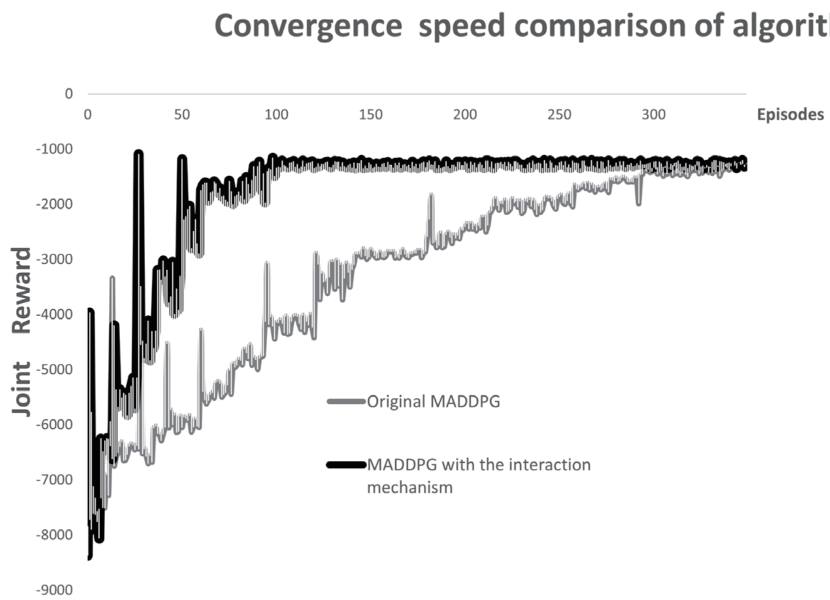

FIGURE 4. Convergence speed comparison of algorithms

value. The curves of reward values with the number of training episodes are performed every 50 episodes. The gray curve represents the original MADDPG algorithm and the black curve represents the MADDPG with interaction mechanism. Since the total reward is set to a negative value, the final joint decision made by multi-type sorting resource agents tends to minimize the reward. Comparing the two curves, it can be concluded that the original MADDPG algorithm converges too slowly due to a large number of multi-type sorting resource agents and unstructured interactions. At the same time, the MADDPG algorithm with the interaction mechanism enables faster convergence and a lower joint reward. It is demonstrated that the MADDPG algorithm with the interaction mechanism can obtain action decisions faster even when the number of sorting resource agents increases.

6. **Conclusion and Outlook.** Aiming to minimize the quality loss of fresh agricultural products during the sorting procedure, this study investigates the collaborative allocation

problem of multi-type sorting resources in rural cold chain warehouses. Firstly, we developed an MARLS-MSRCA and introduced multi-type sorting resource agents into the MARLS-MSRCA, which makes the developed MARLS-MSRCA closer to the real-world resource allocation decision in rural cold chain warehouses. Secondly, we proposed a collaborative decision-making approach based on the MADDPG algorithm and designed an interaction mechanism among the agents to generate the collaborative allocation strategy. The designed interaction mechanism among agents not only conforms to the sequence of resource allocation decisions in the real-world sorting lines but also can address the problem of non-convergence of collaborative decision-making caused by undifferentiated interactions. Finally, we demonstrated the superiority of our collaborative decision-making approach over the human experience and original MADDPG algorithm through numerical experiments in terms of economic benefit and convergence speed analysis.

This study initially explores the research direction of applying MARLS to solve the problem of cooperative allocation of sorting resources in rural cold chain warehouses. However, with the increase in the variety and number of sorting resource agents of MARLS-MSRCA, it might be a prohibitive task to collect all state and action information in a critic due to communication bandwidth and memory limitations [29], and the multi-agent cooperation cannot be realized. Other popular techniques of collaborative MARLS, such as automatic value decomposition [30] and sparse communication network [31]. To improve the resource decision speed and increase the number of collaborative resource agents in MARLS-MSRCA, future research should focus on the automatic value decomposition and the sparse communication network.

Acknowledgment. This work is supported by the Natural Science Foundation of China General Project (71971037), the Innovation Centre for Digital Business and Capital Development of Beijing Technology and Business University (SZSK202207), the NSFC Major International (Regional) Joint Research Project (72010107002) and the Major Project of the National Social Science Foundation (22&ZD151). We also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] J. Su, F. Zhang, S. Chen, X. Yu and S. Sid, Path optimization of fresh products logistics distribution under new retail mode, *International Journal of Innovative Computing, Information and Control*, vol.18, no.2, pp.511-523, 2022.

[2] X. Gao, X. Hu, J. Han, X. Huo, Y. Zhu, T. Liu and J. Ruan, A network flow model of regional transportation of e-commerce and analysis on maturity change of fresh fruit, *International Journal of Innovative Computing, Information and Control*, vol.16, no.3, pp.955-972, 2020.

[3] Z. Zhang, Y. Lu and R. Lu, Development and evaluation of an apple infield grading and sorting system, *Postharvest Biology and Technology*, vol.180, 111588, 2021.

[4] J. Chen and M. Dong, Research on shipment consolidation policy for perishable products by considering shipping freshness requirement, *Systems Engineering – Theory and Practice*, vol.38, no.6, pp.2018-2031, 2018.

[5] L. H. Zhu and X. N. Li, Intelligent sorting method for spherical fruits and vegetables based on combined characteristics, *Computer Engineering and Applications*, vol.52, no.5, pp.173-178, 2016.

[6] J. Yu, W. Miao, G. Zhang, K. Li, Y. Shi and L. Liu, Target positioning and sorting strategy of fruit sorting robot based on image processing, *Traitement du Signal*, vol.38, no.3, pp.797-805, 2021.

[7] F. Wang, J. Zheng, X. Tian, X. Tian, J. Wang, L. Niu and W. Feng, An automatic sorting system for fresh white button mushrooms based on image processing, *Computers and Electronics in Agriculture*, vol.151, pp.416-425, 2018.

[8] K. H. Leung, H. C. W. Lau, D. Nakandala, X. T. R. Kong and G. T. S. Ho, Standardising fresh produce selection and grading process for improving quality assurance in perishable food supply chains: An integrated Fuzzy AHP-TOPSIS framework, *Enterprise Information Systems*, vol.15, no.5, pp.651-675, 2021.

[9] W. Lambrechts, J. S. Klaver, L. Koudijzer and J. Semeijn, Human factors influencing the implementation of cobots in high volume distribution centres, *Logistics*, vol.5, no.32, 2021.

[10] N. N. Yu, T. K. Li and B. L. Wang, Multi-AGV online collaborative scheduling algorithm in automated sorting warehouse, *Computer Integrated Manufacturing Systems*, pp.1-23, 2021.

[11] M. Jiang and G. Q. Huang, Intralogistics synchronization in robotic forward-reserve warehouses for e-commerce last-mile delivery, *Transportation Research Part E: Logistics and Transportation Review*, vol.158, 102619, 2022.

[12] P. Fager, F. Sgarbossa and M. Calzavara, Cost modelling of onboard cobot-supported item sorting in a picking system, *International Journal of Production Research*, vol.59, no.11, pp.3269-3284, 2021.

[13] C. Zhou, A. Stephen, X. Cao and S. Wang, A data-driven business intelligence system for large-scale semi-automated logistics facilities, *International Journal of Production Research*, vol.59, no.8, pp.2250-2268, 2021.

[14] D. Zeng, Z. Zeng and J. Zhan, Evolutionary job scheduling algorithm based on population optimization by deep reinforcement learning, *Application Research of Computers*, vol.39, no.3, pp.739-743, 2022.

[15] H. Tong, Y. Xie, M. Ji and W. Xiang, Online schdeuling of operating room based on multi-agent system, *Computer Applications and Software*, vol.35, no.12, pp.118-122, 2018.

[16] J. Gong, W. Wang, Y. Zhang and Y. Lan, Task assignment strategy of heterogeneous agricultural agent groups based on dynamic stimulus response model, *Transactions of the Chinese Society of Agricultural Machinery*, vol.52, no.5, pp.142-150, 2021.

[17] W. Du and S. Ding, Overview on multi-agent reinforcement learning, *Computer Science*, vol.46, no.8, pp.1-8, 2019.

[18] K. K. Nguyen, T. Q. Duong, G. A. Vien, N. A. Le-Knac and A. N. Nguyen, Non-cooperative energy efficient power allocation game in D2D communication: A multi-agent deep reinforcement learning approach, *IEEE Access*, vol.7, pp.100480-100490, 2019.

[19] H. Zhang, X. Shen, H. Mu, A. Liu and H. Wang, Research on online optimal dispatching of residential energy consumption based on multi-agent asynchronous deep reinforcement learning, *Proceedings of the CSEE*, vol.40, no.1, pp.117-127, 2020.

[20] C. Ying, A. H. F. Chow, H. T. M. Nguyen and K. S. Chin, Multi-agent deep reinforcement learning for adaptive coordinated metro service operations with flexible train composition, *Transportation Research, Part B: Methodological*, vol.161, pp.36-59, 2022.

[21] J. He, J. Zhang, P. Zhang, P. Zheng and M. Wang, Multi-agent reinforcement learning based textile dyeing workshop dynamic scheduling method, *Computer Integrated Manufacturing Systems*, pp.1-31, 2021.

[22] R. Liu, R. Piplani and C. Toro, Deep reinforcement learning for dynamic scheduling of a flexible job shop, *International Journal of Production Research*, vol.60, no.13, pp.4049-4069, 2022.

[23] R. Lowe, Y. Wu, A. Tamar, J. Harb, P. Abbeel and I. Mordatch, Multi-agent actor-critic for mixed cooperative-competitive environments, *Proceedings of Conference on Neural Information Processing Systems*, 2017.

[24] T. P. H. J. Lillicrap, Continuous control with deep reinforcement learning, *International Conference on Learning Representations*, 2016.

[25] V. Konda and J. Tsitsiklis, Actor-Critic algorithms, *Advances in Neural Information Processing Systems*, 1999.

[26] V. K. S. D. Mnih, Human-level control through deep reinforcement learning, *Nature*, vol.7540, no.518, pp.529-533, 2015.

[27] H. Peng and X. Shen, Multi-agent reinforcement learning based resource management in MEC- and UAV-assisted vehicular networks, *IEEE Journal on Selected Areas in Communications*, vol.39, no.1, pp.131-141, 2021.

[28] *Baishui County People's Government-Analysis of Apple Production Situation in 2021 in Whitewater*, http://www.baishui.gov.cn/xwzx/bmdt/83376.htm, 2021.

[29] A. Oroojlooy and D. Hajinezhad, A review of cooperative multi-agent deep reinforcement learning, *Applied Intelligence*, 2022.

[30] P. Sunehag, L. Guy, G. Audrunas, M. Wojciech, Z. Vinicius, J. Max, L. Marc et al., Value-decomposition networks for cooperative multi-agent learning based on team reward, *Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'18)*, pp.2085-2087, 2018.

[31] L. Cassano, K. Yuan and A. H. Sayed, Multiagent fully decentralized value function learning with linear convergence rates, *IEEE Transactions on Automatic Control*, vol.66, no.4, pp.1497-1512, 2021.

Author Biography

Di Wang received her B.S. degree from the School of Economics and Management, Hainan University, China, in 2018 and her M.S. degree from the Discipline of Logistics Engineering, Dalian University of Technology, Dalian, China, in 2020, respectively. She is currently working towards the Ph.D. degree in Dalian University of Technology, Dalian, China. Her research interests include collaborative optimization of cold chain sorting resources and multi-agent reinforcement learning system.

Haiyang Shi received his B.S. degree from the School of Management, Northwest Polytechnic University, Xi'an, China, in 2014 and his M.S. degree from the Discipline of Management Science and Engineering, Dalian University of Technology, Dalian, China, in 2017, respectively. He is currently working towards the Ph.D. degree in Dalian University of Technology, Dalian, China. His research interests include intelligent decision-making in E-commerce logistics and multi-objective optimization.

Lijun Sun received her M.S., and Ph.D. degrees from the Department of Management Science and Engineering, Dalian University of Technology, China, in 2007 and 2011, respectively. She is now working as a Professor in the School of Economics and Management, Dalian University of Technology. Her areas of interests include intelligent decision-making and optimizations in logistics system. She has published over 40 journal papers, including some published in high quality journals, like Transportation Research Part E, Decision Support Systems, and International Journal of Production Research.

Yutian Du receives her B.S. degree from Jilin University. She is a Master student at School of Economics and Management, Dalian University of Technology, Dalian, China. Her research interest focuses on data-driven approaches to multistage stochastic linear optimization.