Optimization of Fourth Party Logistics Routing Considering Infection Risk and Delay Risk

Guihua Bo[1], Sijia Li[2,*], Mingqiang Yin[3], Mingkun Chen[4], and Xin Liu[5]
School of Information and Control Engineering, Liaoning Petrochemical University, Fushun, China[1, 2, 3, 4, 5]

***Abstract*—In the context of the rapid development of e-commerce and the increasing demands for logistics services, particularly in the face of challenges posed by public health emergencies, this paper explores how to integrate supply chain resources and optimize delivery processes. It provides an in-depth analysis of the characteristics of the Fourth Party Logistics Routing Optimization Problem (4PLROP) in complex environments, specifically focusing on the impacts of infection risk and delay risk, and proposes a new risk measurement tool. By constructing a mathematical model aimed at minimizing Conditional Value-at-Risk (CVaR) and improved Q-learning algorithm, the study addresses the 4PLROP while considering cost and risk constraints. This approach enhances the efficiency and service quality of the logistics industry, offers effective strategies for 4PL companies in the face of uncertainty, and provides customers with safer and more reliable logistics solutions, contributing to sustainable development.**

***Keywords*—Logistics services; public health emergencies; logistics routing optimization; improved Q-learning algorithm; CVaR; infection risk**

I. INTRODUCTION

As the limitations of Third Party Logistics (3PL) capabilities become increasingly apparent, the traditional route planning problem is transitioning into the more complex Fourth Party Logistics Routing Optimization Problem (4PLROP). 4PL, known as the integrator of supply chains, consolidates its own resources, capabilities, and technologies, as well as those of other 3PL providers, to offer comprehensive supply chain solutions to clients. The concept of 4PL has garnered widespread attention from both the industry and academia since its inception. Presently, 4PL enterprises or platforms such as UPS, Cainiao, and Ningbo Fourth Party Logistics Market have established long-term cooperative relationships with manufacturing enterprises like Haier, providing them with specialized logistics and transportation services. In the academic sphere, numerous scholars have conducted research on various aspects of 4PL, including route optimization [1], network design [2], risk management [3], combinatorial auctions [4]-[5], supply chain integration [6], and information technology application [7]. Route optimization in 4PL, as one of the core issues at the tactical layer of 4PL operation and management, has been receiving considerable scholarly attention in recent years. It integrates 3PL and route selection decisions to enhance the efficiency of logistics transportation [8].

In the advent of public health emergencies, the logistics industry has played a pivotal role in ensuring resource supply, yet such occurrences also introduce new challenges and risks to route planning. During the 2020 pandemic, international and Hong Kong, Macau, and Taiwan air passenger volumes plummeted by over 15%, particularly in key logistics hubs like Wuhan, where containment measures significantly impacted logistics routes. To circumvent high-risk pandemic areas, logistics enterprises were compelled to rechart transportation routes, which not only heightened the complexity of route planning but also augmented the time required. These shifts demand that logistics enterprises factor in the risks posed by pandemics during the planning process. 4PL must consider the risks brought about by pandemics and other public health emergencies, employing agile supply chain management and efficient resource allocation to mitigate the impact of these risks on the logistics network. In 4PL, such risk management is especially crucial. As the coordinator and integrator within the supply chain, 4PL is tasked with managing the demands and risks of multiple clients while ensuring service quality.

This article delves into the 4PLROP, which takes into account the risks of infection and delays, against the backdrop of the pandemic. Cities are categorized based on risk levels, and the infection risks of various cities are assessed using quantitative methods. Leveraging the extensive application of Conditional Value-at-Risk (CVaR) in the optimization domain, a mathematical model is established with the objective of minimizing CVaR, and constraints are set for delivery costs and infection risks. An improved Q-learning algorithm is employed to solve this model, aiming to identify the route that satisfies the conditions of the lowest CVaR for both infection risks and delivery costs. In this manner, 4PL can better adapt to the ever-changing market environment, ensuring the stability and efficiency of the supply chain.

The principal contributions of this article are as follows:

- In the context of public health emergencies, a novel 4PL route optimization problem that concurrently considers the risks of infection and delays has been investigated;
- The CVaR metric is employed to characterize risks, and a nonlinear programming model is established with constraints on delivery costs and infection risks, aiming to minimize the risk as the objective;
- An improved Q-learning algorithm is proposed to solve the model presented. Through this approach, 4PL can better adapt to the ever-changing market environment, ensuring the stability and efficiency of the supply chain.

The remainder of this article is structured as follows:

The remainder of this article is structured as follows: Section II provides a review of literature pertaining to 4PL route-related studies. Section III delineates the problem and elucidates the model and notation employed. Section IV applies the Q-learning algorithm to the 4PL route optimization problem, where the optimal path planning is achieved through the establishment of action-state pairs, construction of a reward function, enhancement of exploration strategies, and model training. Section V validates the efficacy of the improved Q-learning algorithm in the context of 4PL route optimization through experimental analysis, demonstrating the algorithm's high solution speed and stability across various scales of test cases, and its ability to provide customers with delivery routes that minimize risk at a specific confidence level. Section VI presents our conclusions and prospective directions for future research.

II. Literature Review

In the context of public health emergencies, a plethora of theoretical foundations and practical case studies has been provided by existing research to address the 4PLROP. Within this section, an exhaustive review of the pertinent literature on 4PLROP has been conducted. The existing research delineates the complexities and challenges posed by the emergence of public health crises on 4PLROP, highlighting the need for innovative approaches to mitigate the associated risks and delays.

In the field of 4PLROP, a relatively early study dates back to 1998. The concept of 4PL was initially introduced by Andersen Consulting [9]. Since then, based on this concept, a plethora of research on 4PLROP has been conducted: Huang et al. [10] conducted research on the 4PLROP with uncertain delivery time in emergencies. Huang et al. [11] proposed an improved genetic algorithm based on simple graphs and the Dijkstra algorithm to preclude the emergence of infeasible solutions in 4PLROP. Ren et al. [12] designed a genetic algorithm embedded with the Dijkstra algorithm to solve the 4PLROP problem, thereby laying the foundation for the research on 4PLROP. The existing studies can generally be categorized into two types. The first type is the 4PL route planning problem in a deterministic environment, and the second type is the 4PLROP in an uncertain environment.

In the realm of deterministic environment, an early scholar established a directed graph model to optimize the selection of routes, transportation modes, and third-party logistics providers [13]. Subsequently, a 4PL optimization model was constructed by another scholar to streamline the corresponding 4PLROP [14]. Thereafter, a 4PLROP approach based on the immune algorithm was proposed by some scholars, enhancing the algorithm's capability to address 4PLROP [15]. Building on this foundation, a mathematical model for point-to-point multi-task 4PLROP without edge repetition was established by other scholars, considering the cost and time attributes of each node and edge, and an ant colony optimization algorithm was designed to solve the path optimization problem [16]. Recently, Zhou et al. [17] addressed the 4PLROP problem considering cost discounts by minimizing operational costs, taking into account customer delivery deadlines and transportation capacity constraints. Cai et al. [15] minimized the linear combination of transportation and time costs by considering certain customer preference factors.

In the realm of uncertainty, Huang et al. [18] proposed an uncertain programming model for the 4PLROP in emergency situations. The effectiveness of this model was verified through comparison with the stochastic programming model and numerical experiments. Huang et al. [19] transformed the uncertainty theory into a deterministic model and designed an improved genetic algorithm for solving to address the 4PLROP in an uncertain environment. Lu et al. [20] solved the uncertain delivery time control model of 4PLROP through the genetic algorithm. Lu et al. [21] dealt with the 4PLROP problem under the conditions of uncertainty in 3PL transportation time, transportation cost, node transfer time and transfer cost, and designed a solution model using the grey wolf optimization algorithm improved by the ant colony system. Lu et al. [22] considered the uncertainties in transportation time and cost caused by seasonal and human factors, constructed a multi-objective chance-constrained programming model aiming to minimize transportation time and cost, and proposed a hybrid beetle swarm optimization algorithm combined with the Dijkstra algorithm to solve the problem. Ren et al. [23] established a 4PLROP chance model with time windows and random transportation time under the constraint of total transportation cost, aiming to maximize the chance that the total transportation time meets the time windows. And the ant colony algorithm was used to solve the deterministic model. Gao et al. [24] applied the uncertain stochastic programming model to solve the 4PLROP with random demand and uncertain transportation and transshipment times, aiming to minimize the total transportation cost under various constraints. Recently, Ren et al. [25] aimed to examine the impact of decision-makers' risk preferences on the 4PLROP, contributing to the analysis of logistics behavior and route integration optimization in uncertain environments.

In addition to the aforementioned studies, recent years have witnessed a growing focus on the risk factors in 4PLROP. Deng et al. [26] utilized the ant-colony algorithm to address the mathematical model of 4PLROP, where the Value at Risk (VaR) was employed to represent the delay risk in an uncertain environment. Bo et al. [27] carried out research on the 4PLROP with tardiness risk by introducing VaR to measure the time-related risk. Wang et al. [28] established a mathematical model considering customers' risk-averse behavior and studied the 4PLROP in the context of customers' risk-avoidance behavior. Recently, Liu et al. [29] introduced the risk value VaR to measure the risk of delays, which has been a significant advancement in the risk assessment of 4PLROP.

The probability that the delay quantity is less than a certain value, as denoted by VaR, is required to be greater than or equal to the confidence level prescribed by the client. However, it merely takes into account the likelihood of the occurrence of delay risks, without considering the mean of such risks when they materialize under extreme conditions. The conditional mean of delay risks exceeding VaR can be determined by the CVaR model. By integrating the risk level of the distribution plan and the anticipated delay risks, rational decisions concerning distribution services can be formulated.

In summary, while the existing literature encompasses a multitude of aspects of 4PLROP, there is still a deficiency in addressing the infection risks and delays associated with public health emergencies. This article provides a novel perspective and practical methodologies for this field of research by constructing corresponding mathematical models based on CVaR and employing an improved Q-learning algorithm. The aim is to assist 4PL systems in better adapting to the ever-changing market environment, ensuring the stability and efficiency of the supply chain.

III. Problem Description and Model Assumptions

A. The Path Optimization Problem in the Context of the Pandemic

In environments where logistics warehouses and transfer node cities are densely populated with personnel and abundant goods, the potential risk of virus transmission during the pandemic cannot be overlooked. Especially in the context of ongoing pandemic prevention and control, 4PL service providers must ensure that only goods, and not viruses, are transported by vehicles while maintaining smooth logistics operations. Consequently, stringent monitoring of infection risks in logistics transportation has become a crucial component of epidemic prevention efforts.

During the special period of the pandemic, to avoid potential infections in high-risk areas, this paper has developed a specific planning strategy for delivery routes, incorporating infection risk constraints. This issue can be specifically described as follows: as graphical illustration of the process in Fig. 1, a multi-layer graph "G=(V,E)" is used to represent the 4PLROP, where "|V|=n" is the set of node cities and E is the set of edges. The node city s represents the supply city, node city t represents the destination city, and other node cities representing transfer node cities. The number of node cities "|V|=n" indicates the total number of node cities, each of which has attributes such as time, cost, carrying capacity, and reputation. Since there may be multiple 3PL suppliers offering services between any two node cities, there are multiple edges between any two node cities in the graph (each edge represents a different 3PL, identified by a unique number). Consequently, each edge has different attributes related to time, cost, capacity, and reputation, meaning that each 3PL has its corresponding properties. Therefore, when selecting transfer node cities and 3PL suppliers, it is necessary to comprehensively weigh various factors to ensure that the chosen path is not only cost-effective and time-efficient but also minimizes infection risks to the greatest extent possible.

The goal is to provide customers with a delivery plan that meets cost budget and infection risk control requirements while minimizing CVaR, ensuring that goods are delivered safely and on time. Therefore, the following assumptions are proposed to address the aforementioned research issues:

- (1): It is assumed that infection risks only exist at node cities where 3PL suppliers are changed and where handling and unloading occur, while the transportation between node cities is considered risk-free.
- (2): It is assumed that the level of infection risk is directly related to the cumulative number of locally confirmed cases in the city over the past 14 days, and that high-risk node cities are strictly avoided. The specific risk assessment method will be detailed in Section III (C).

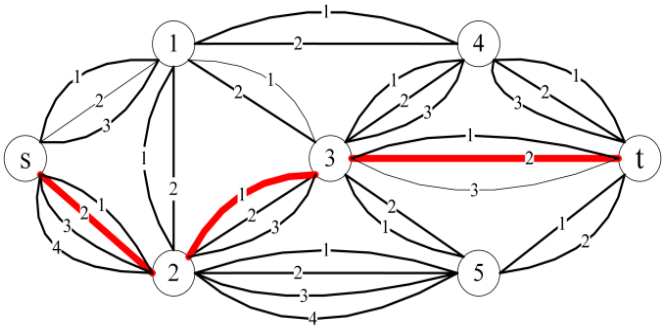

Fig. 1. 7-node problem description.

B. Parameters and Variables

By defining the following parameters and variables to establish a mathematical model, as shown in Table I.

TABLE I. Definitions of Parameters and Variables

Symbol	Definition description
r_{ij}	The number of 3PL providers that can offer delivery services between the node city *i* and node city *j* (namely, the number of edges between the two node cities).
C_{ijk}	The transportation cost required by the k-th 3PL provider for delivery services between the node city *i* and node city *j*.
T_{ijk}	The random transportation time required by the k-th 3PL provider for services between the node city *i* and node city *j*.
$C_j^{'}$	The transshipment cost required when passing through node city *j*.
$T_j^{'}$	The random transshipment time required when passing through node city *j*.
R	The set of node cities and edges contained in the path is, namely, $R=\{v_s,\cdots,v_i,k,v_j,\cdots,v_t\}$. As shown in Fig. 1, the red path can be represented by $R=\{v_s,2,v_2,1,v_3,2,v_t\}$.
$x_{ijk}(R)$	Decision variable. When the 3PL provider represented by the k-th edge between city *i* and node city *j* provides the distribution task, it takes 1; otherwise, it takes 0. As shown in (1).
$y_j(R)$	Decision variable. If the city represented by node city *j* provides the transshipment task, it takes the value of 1; otherwise, it takes 0. As shown in (2).
X_j	The cumulative number of local confirmed cases within 14 days in node city *j*.
f	The unit person-time infection risk of the cumulative number of local confirmed cases within 14 days.
F_0	The maximum acceptable infection risk for customers.

$$x_{ijk}=\begin{cases}1, & \text{The k-th edge between i and j}\\ & \text{belong to path } R\\ 0, & \text{else}\end{cases} \quad (1)$$

$$y_j(R)=\begin{cases}1, & \text{node city j belong to path } R\\ 0, & \text{else}\end{cases} \quad (2)$$

C. Quantification of the Infection Risk

This paper studies the 4PLROP during the early stages of the pandemic. Therefore, it draws on the classification of cities into low, medium, and high-risk areas established in the early phase of the pandemic to assign infection risk values to the node cities, as shown in Table II.

TABLE II. RISK CLASSIFICATION CRITERIA

Risk rating	Classification criterion	Response policies
Low-risk area	Within 14 days, without any new or existing confirmed cases.	Strengthen external prevention and control, fully restart production and daily life, and lift road traffic restrictions
Medium-risk area	Within 14 days, if the number of new confirmed cases is ≤ 50 or there are no cluster outbreaks, even if the cumulative confirmed cases exceed 50.	Implement a dual prevention and control strategy, steadily and orderly restore the normal state of production and daily life
High-risk area	Cumulative cases exceed 50, and there have been cluster outbreaks in the past 14 days.	Implement strict management with dual-direction prevention and control, ensuring that the pandemic does not spread or overflow

According to the defined standards, when assessing the COVID-19 infection risk in a certain area, the number of locally confirmed cases over a continuous fourteen-day period is considered. If an area has no locally confirmed cases or has no new cases for fourteen consecutive days, it is regarded as low-risk. If there are new cases within fourteen days but the cumulative confirmed cases do not exceed 50, it is classified as a medium-risk area. When the cumulative confirmed cases exceed 50, the area is considered high-risk. Given that the incubation period of the COVID-19 virus is fourteen days and that travel codes also reference the travel history within the last fourteen days, this paper uses the number of locally confirmed cases in a node city over the past fourteen days as the basis. The infection risk f is quantified in terms of the number of confirmed cases per person. For example, if the cumulative confirmed cases in a node city within fourteen days amount to 25, then the infection risk is 25f. Moreover, infection risk occurs only during the transfer at the node city.

D. Mathematical Model

Under the confidence level β, when minimizing CVaR, calculate the distribution path with the minimum average overdue risk. Add the infection risk constraint and establish the following mathematical model:

$$min\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k}^{r_{ij}}\mu_{ijk}x_{ijk}(R)+\sum_{j=1}^{n}\mu_j y_j-T_0\right)+c_1(\beta)\times\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k}^{r_{ij}}\delta_{ijk}^2x_{ijk}^2(R)+\sum_{j=1}^{n}\delta_j^2y_j^2(R)} \quad (3)$$

$$s.t. \quad \sum_{j=1}^{n}X_j f y_j \le F_0 \quad (4)$$

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{r_{ij}}C_{ijk}x_{ijk}(R)+\sum_{j=1}^{n}C_j'y_j(R)\le C_0 \quad (5)$$

$$\Delta T=\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{r_{ij}}T_{ijk}x_{ijk}(R)+\sum_{j=1}^{n}T_j'y_j(R)-T_0 \quad (6)$$

$$R=\{v_s,\cdots,v_i,k,v_j,\cdots,v_k\}\in G \quad (7)$$

$$x_{ijk}(R),y_j(R)\in\{0,1\} \quad (8)$$

$$X_j<50 \quad (9)$$

Among them, Constraints (4) the capacity limit for the infection risk, F_0 denotes the maximum acceptable infection risk given by the customer. Constraints (5) the capacity limit for the delivery cost, where C_0 is the maximum cost acceptable to the customer. Constraints (6) is the expression of the overdue quantity ΔT, which is a random variable. Constraints (7) reflects the path to ensure that the path is a legal connected path from the initial node city to the destination city. Constraints (8) manifests $x_{ijk}(R)$ and $y_j(R)$ are decision variables. Constraints (9) conveys that the transportation path cannot pass through high-risk areas.

IV. ALGORITHM DESIGN

When the improved Q-learning algorithm is employed to address the 4PLROP, the primary procedures encompass the initialization of parameters, the establishment of action-state settings, the construction of the reward function, the formulation of exploration strategies and the training of the model.

A. Action-state Setting

This paper combines improved Q-learning with the 4PLROP, viewing the choice of actions as related to the selection of 3PL suppliers and treating node cities as different states. Let s represent the current state (the current node city). The corresponding 3PL suppliers for this node city can be represented by the action space $A=\{a_1,a_2,\cdots,a_k,\cdots,a_K\}$,k=1,2, ⋯, K . Each action-state pair corresponds to a Q-value.

Taking the 7-node for example, refer to Fig. 1 for the illustration, the initial node city can be regarded as the initial node city s. The node cities connected to the initial node city are node city 1 and node city 2. There are three selectable 3PL suppliers corresponding to the route from the initial node city to node city 1, labeled as 1, 2, and 3, which can be designated as actions α_1,α_2 and α_3. Similarly, there are four selectable 3PL suppliers for the route from the initial node city to node city 2, labeled as 1, 2, 3, and 4, which can be designated as actions $\alpha_4,\alpha_5,\alpha_6,\alpha_7$. Thus, there are 7 actions available at the initial node city. When choosing actions α_1,α_2 and α_3 at the initial node city, the next state transitions to node city 1. Likewise, when choosing actions $\alpha_4,\alpha_5,\alpha_6,\alpha_7$ at the initial node city, the next state transitions to node city 2. The action sets for the other node cities can be defined in a similar manner. Using the 7-node as an example, the selected actions and their corresponding next states are shown in Table III.

TABLE III. 7-NODE PROBLEM ACTIONS AND CORRESPONDING NEXT STATES SETTINGS

The selected actions	Corresponding to the next state
$A=\{a_1,a_2,a_3\}$	Transfer node city 1
$A=\{a_4,a_5,a_9\}$	Transfer node city 2
$A=\{a_{10},a_{11},a_{14},a_{15},a_{16}\}$	Transfer node city 3
$A=\{a_{12},a_{13},a_{21},a_{22},a_{23}\}$	Transfer node city 4
$A=\{a_{17},a_{18},a_{19},a_{20},a_{24},a_{25}\}$	Transfer node city 5
$A=\{a_{26},a_{27},a_{33}\}$	Demand node city t

B. The Construction of the Reward Function

Since the Q-learning algorithm is based on the Markov Decision Process (MDP) model, a discrete reward and punishment function is adopted for computational convenience [30]. Given that transportation tasks cannot pass through high-risk areas, the number of infections at the transfer node cities along the path must not exceed 50. Therefore, when the number of infections in a certain city *j* is less than or equal to 50, the reward is 1. Conversely, when the number of infections in city *j* exceeds 50, the reward is -100, as depicted in Eq. (10).

$$r_0(s,a)=\begin{cases}1 & if \quad X_j \le 50\\ -100 & if \quad X_j > 50\end{cases} \tag{10}$$

When there is a connection between node city *i* and node city *j* and *j* is not the destination, the reward is 1. When there is no connection between node city *i* and node city *j*, the reward is -1. When there is a connection between node city *i* and node city *j* and *j* is the destination, the reward is 100, in (11) illustrates this concept.

$$f(x)=\begin{cases} 1\text{ , i, j are connectedj is not the end node city;}\\ \text{-1,} \qquad \text{i, j are not connected ;}\\ \text{100, i, j are connected;j is not the end node city}\end{cases} \tag{11}$$

Considering the magnitude of rewards is related to the mean and variance within the objective function, and also needs to meet certain constraints, the reward function for this issue can therefore be rewritten as shown in study (12). The parameter ω_1 is inversely proportional to the distribution cost corresponding to the selected 3PL supplier and transfer node city, that is, the smaller the distribution cost, the greater the reward value obtained, as illustrated in study (13). In a parallel manner, ω_2 is inversely proportional to the mean value of the distribution time corresponding to the selected 3PL supplier and transfer node city. When the mean value of the random time is smaller, the greater the reward value obtained, as evidenced in study (14). The parameter ω_3 is inversely related to the average distribution time associated with the selected 3PL supplier and transfer node city, When the variance is smaller, the corresponding reward value is greater, where *k1* and k2 are the weighting coefficients of the reward function, as detailed in study (15). Additionally, ω_4 is correlated with the infection count at the node city, as elucidated in study (16).

$$r=\omega_1 r(s,a)+\omega_2 r(s,a)+\omega_3 r(s,a)+\omega_4 r_0(s,a) \tag{12}$$

$$\omega_1=\frac{k_1}{C_{ijk}+C_j} \tag{13}$$

$$\omega_2=\frac{k_2}{\mu_{ijk}+\mu_j} \tag{14}$$

$$\omega_3=\frac{1-k_1-k_2}{\delta_{ijk}^2+\delta_j^2} \tag{15}$$

$$\omega_4=\frac{1}{X_j} \tag{16}$$

C. The Exploration Strategy of Improved Q-learning Algorithm

When the agent interacts with the environment to learn, it must choose known actions that maximize the reward while also ensuring that it can learn more experiences in an unknown environment, thereby laying the foundation for obtaining more cumulative rewards. Therefore, it is essential to establish an appropriate exploration strategy to achieve optimal training results. The traditional Q-learning algorithm typically employs the ε-greedy strategy as its exploration method.

The mathematical description of the *ε*-greedy strategy is as follows:

$$\pi(a,s)=\begin{cases}arg\,max\,Q(s,a) & 1-\varepsilon\\ q_{random} & \varepsilon\end{cases} \tag{17}$$

Eq. (17), it can be understood as randomly selecting the selectable actions in the current state with a certain probability *ε*, and choosing the action corresponding to the maximum Q value among the current actions with a probability of *1-ε*.

When using the Q-learning algorithm to address the 4PLROP, the environment is relatively simple, and both states and actions are limited. Therefore, it is necessary to establish a corresponding reinforcement learning environment based on the characteristics of the problem. To better explore the environment, this paper adopts a random strategy more suitable for the problem to select actions. According to the reward matrix established by the reward function, it is observed that in the current state, when the reward value is -1, the two node cities are disconnected. Therefore, the exploration strategy is set to randomly select actions with reward values greater than -1 in the current state to reduce exploration time.

D. Model Training

By designing a Q-table to train the agent, each row in the Q-table represents all the states available to the agent, while each column represents the actions the agent can perform in the corresponding state. Each state in the multi-layer graph represents different node cities, and each action represents different 3PL suppliers in the multi-layer graph. Initially, all states in the Q-table are set to 0. The reward values obtained from executing different actions (selecting different suppliers) are then calculated based on the reward matrix established by the reward function, and the values of the elements in the Q-table are updated using Eq. (18). Each iteration is considered a training session for the agent. During each training session, the agent attempts to move from the initial node city to the destination node city, updating the elements in the Q-table after executing each action.

$$Q^*(s,a) \leftarrow Q(s,a) + q\left[R(s,a,s) + \gamma \cdot \max Q(s,a) - Q(s,a)\right] \quad (18)$$

E. The Flow Chart of Q-learning Algorithm

When using the improved Q-learning algorithm to solve the 4PLROP, firstly, based on the existing data, the elements in the matrix are initialized by using the reward function. Since there are multiple different 3PL suppliers between two node cities, that is, one state corresponds to multiple ones. Therefore, it is necessary to set the actions corresponding to each state, and then train and update the matrix Q through the setting of the matrix *R* and related parameters. Finally, the optimal path planning can be obtained based on the Q-table. The flowchart of the 4PLROP using the improved Q-learning algorithm is shown in Fig. 2. The specific steps are as follows:

Step 1: Load the known data information in MATLAB.

Step 2: Initialize the parameters γ、α and the Q-table, set the initial state and the final state, and at the same time, use the given data Eq. (12) to construct the reward matrix R.

Step 3: Set the initial state as the initial node city.

Step 4: Determine the action through the random selection strategy, that is, select a feasible 3PL supplier.

Step 5: Perform the action α (namely, select a 3PL supplier of the current node city), and then transfer to the new state s'(node city). Update the Q-table based on the reward matrix R and preset parameters.

Step 6: Determine whether s' is the final node city. If not, return to Step 4; if yes, proceed to Step 7.

Step 7: Determine whether the set number of training times has been completed. If not, return to Step 3 to continue training; if yes, proceed to Step 8.

Step 8: The training process is over and the final Q-table is output.

Step 9: Combine the Q-table to determine and output the best logistics distribution plan.

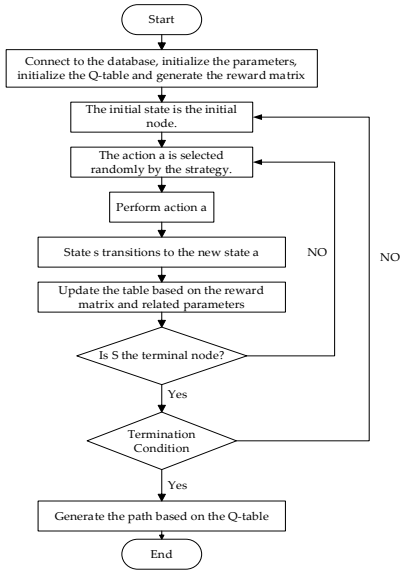

Fig. 2. Flow chart of improved Q-learning algorithm.

V. Experimental Results and Analysis

Referring to Fig. 3, it is the epidemic data chart for some periods in 2022. Fig. 4 illustrates the cumulative number of confirmed cases in some cities across the country within 14 days. These are used as the data references for this section. The relevant data comes from the National Health Commission of China and the health commissions of various provinces and cities.

This section first takes the 7-node as an example to analyze the influence of the training times episode, discount factor γ, and learning rate α in the Q-learning algorithm on the calculation results, and obtains a set of optimal parameter combinations. Then, it solves the mathematical model with the constraint conditions of the delivery cost and the infection risk and the objective function of minimizing CVaR. Finally, the best distribution path obtained from the corresponding example is visualized. The software used by the algorithm is MATLAB 2023a, and the operating environment is Intel(R) Core(TM) i7-2600 @3.40GHz.

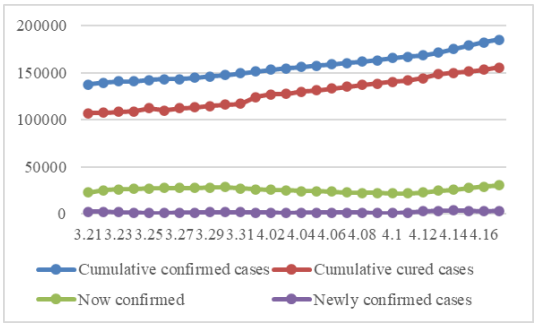

Fig. 3. Epidemic data chart for partial periods in 2022.

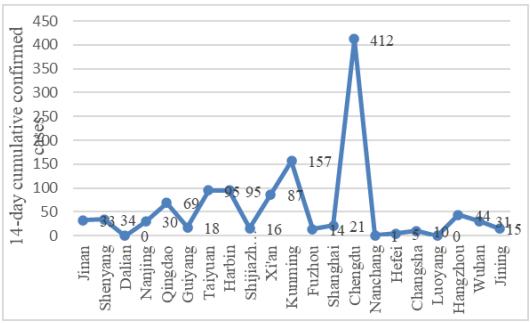

Fig. 4. Cumulative number of confirmed cases within 14 days in some cities.

A. Parameter Test

The parameters within the improved Q-learning algorithm are rigorously tested through extensive experimental simulations. This is achieved by keeping all other parameters constant and observing the impact of variations in a single parameter on the solution outcomes. The efficacy is denoted by the optimality rate, which represents the probability of obtaining the best solution during the execution of the algorithm.

Through repeated experiments on *k1* and *k2* in the reward function in different sized examples, when the values of *k1* and *k2* are the data in Table IV, the algorithm has the best solution effect.

TABLE IV. PARAMETER SETTINGS FOR DIFFERENT INSTANCES

Number of node	k1	k2	eposide	CVaR	Best path	Time
7	0.1	0.8	100	27.7892	$R=\{v_s,2,v_2,2,v_3,1,v_t\}$	0.9s
15	0.1	0.2	200	12.1589	$R=\{v_s,3,v_3,2,v_6,3,v_{13},2,v_t\}$	1s

Fig. 5-7 offer a graphical representation of the data, the parameter test process of the improved Q-learning algorithm for solving the 7-node example is presented when the confidence level is 0.9, the infection risk constraint is 8.5×10-5, and the cost constraint is 80. Wherein the " best rate" refers to the probability that the best solution is achieved during the algorithm's execution, with the total number of runs set to 100.The test results show that the best parameters of the improved Q-learning algorithm are γ=0.8, α=0.9 and episode = 100, respectively.

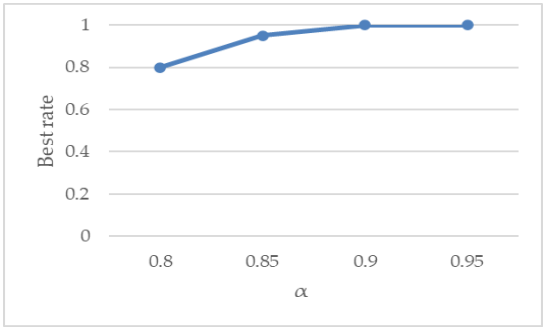

Fig. 5. Parameters α performance analysis.

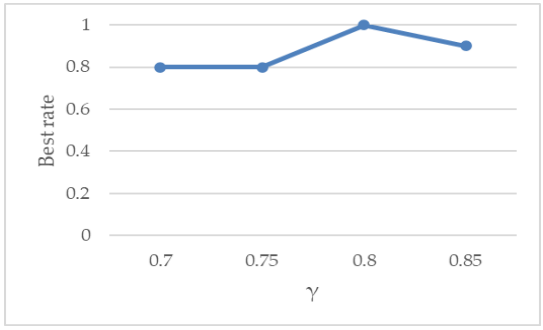

Fig. 6. Parameters γ performance analysis.

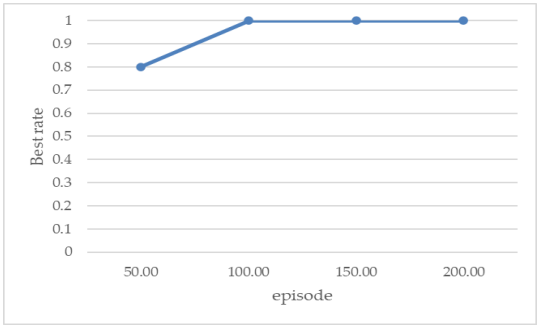

Fig. 7. Parameters episode performance analysis.

B. Case Analysis

To validate the effectiveness of the proposed model, we solved two instances of different scales, with 7-node and 15-node, and obtained the corresponding minimum CVaR values and the best routes. The information related to the 7-node and edges is shown in Table V and Table VI. Due to the large amount of information from the 3PL suppliers, only partial information is provided in Table VI.

TABLE V. 7-NODE CALCULATION EXAMPLE RELATED INFORMATION

Node	Cost	The mean of random time	The variance of random time	The number of infections
s	10	6	25	27
1	12	4	36	11
2	6	5	16	23
3	9	7	64	20
4	11	4	9	51
5	15	6	49	35
t	7	5	25	10

TABLE VI. 7-NODE CALCULATION EXAMPLE 3PL SUPPLIER RELATED INFORMATION

Initial	End	3PL Number	Transportation cost	The mean of random time	The variance of random time
s	1	1	20	12	169
s	1	2	18	15	196
s	1	3	24	10	64
s	2	1	18	16	225
s	2	2	17	17	196
s	2	3	19	14	169
s	2	4	15	20	329

As depicted in Fig. 8, it is the path diagram corresponding to the solution result of 7-node. Among them, the black pentagram s represents the initial node city, the black pentagram t is the destination node city, and the node city 4 marked by the red pentagram indicates the transfer node city that does not meet the constraint of the number of infections. That is, transfer node city 4 is in a high-risk area, so the path cannot pass through node city 4. The other transfer node cities that meet the constraint of the number of infections. The path marked by the blue thick line is the distribution path corresponding to the minimum CVaR that satisfies the constraints of cost and the infection risk, and the detailed data is shown in Table VI.

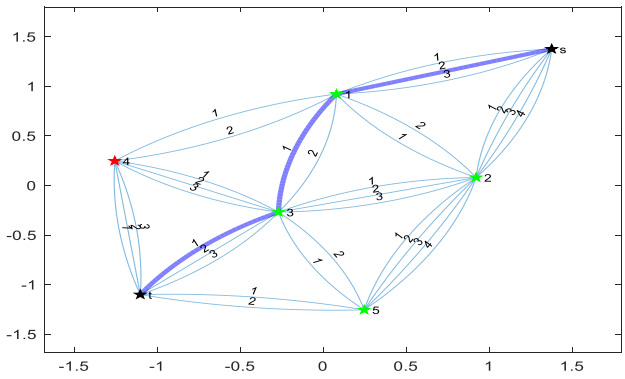

Fig. 8. 7-node solution path diagram.

The information is summarized in Table VII, the solution of the CVaR model of the 7-node problem are given under different confidence levels when the cost constraint C_0=80 and the maximum acceptable the infection risk given by the customer is 8.5×10^{-5}. Among them, "β" represents the confidence level, that is, the degree of risk aversion of the customer, "*CVaR*" is the best solution obtained by the CVaR model (the evaluation criterion is the objective function), "Best path" is the distribution path corresponding to the best solution obtained, "*F*" is the infection risk corresponding to the distribution path of the best solution obtained, "*Best rate*" indicates the probability that the best solution obtained by the algorithm accounts for the total number of runs of the algorithm. At this time, the total number of runs is 100, and "*Time*" is the running time of the algorithm for one run, in seconds.

TABLE VII. SOLUTION OF 7-NODE PROBLEMS AT T_0=70、C_0=80 AND F_0=8.5×10^{-5}

β	eposide	CVaR	Best path	F	Best rate	Time
0.9	100	27.7892	R={v_s,2,v_1, 1,v_3,1,v_t}	7.352×10^{-5}	0.98	0.9s
0.95	100	35.1168	R={v_s,2,v_1, 1,v_3,1,v_t}	7.352×10^{-5}	0.98	0.9s
0.99	100	49.4634	R={v_s,2,v_1, 1,v_3,1,v_t}	7.352×10^{-5}	0.98	0.9s

According to the data in Table VII, when the confidence level is 0.9, the Delivery Cost is C_0=80, and the infection risk constraint is F_0=8.5×10^{-5}, the obtained optimal CVaR value is 27.7892, indicating that the corresponding average delay risk of the distribution task is 27.7892, the corresponding distribution cost is 80, the infection risk is 7.352×10^{-5}, and the corresponding best distribution path is $R=\{v_s,2,v_1,1,v_3,1,v_t\}$, indicating that when transporting from the source node city s to the destination node city t, the selected transfer node cities are 1 and 3 respectively, and the 3PL supplier number selected between each two transfer node cities is 2, 1, 1; when the confidence level is 0.95, the cost constraint is C_0=80, and the infection risk constraint is F_0=8.5×10^{-5}, the obtained minimum CVaR value is 35.1168, indicating that the corresponding average delay risk of the distribution task is 35.1168, the corresponding distribution cost is 80, the infection risk is 7.352×10^{-5}, and the corresponding best distribution path is $R=\{v_s,2,v_1,1,v_3,1,v_t\}$, indicating that when transporting from the source node city s to the destination node city t, the selected transfer node cities are 1 and 3 respectively, and the 3PL supplier number selected between each two transfer node cities is 2, 1, 1; when the confidence level is 0.99, the cost constraint is C_0=80, and the infection risk constraint is F_0=8.5×10^{-5}, the obtained optimal CVaR value is 49.4634, indicating that the corresponding average delay risk of the distribution task is 49.4634, the corresponding distribution cost is 80, the infection risk is 7.352×10^{-5}, and the corresponding best distribution path is $R=\{v_s,2,v_1,1,v_3,1,v_t\}$, indicating that when transporting from the source node city *s* to the destination node city *t*, the selected transfer node cities are 1 and 3 respectively, and the 3PL supplier number selected between each two node cities is 2, 1, 1.

The relevant information of 15-node is shown in Table VIII. Since there are 91 rows of information corresponding to the 3PL supplier of 15-node, only a partial information is displayed in Table IX.

TABLE VIII. 15-NODE CALCULATION EXAMPLE RELATED INFORMATION

Node	Cost	The mean of random time	The variance of random time	The number of infections
s	10	6	4	34
1	12	7	4	33
2	8	4	1	95
3	6	5	1	16
4	14	8	4	87
5	9	6	4	30
6	8	5	1	15
7	12	6	4	69
8	10	5	1	31
9	11	6	4	44
10	9	6	1	18
11	14	7	4	21
12	8	5	1	10
13	15	6	4	14
t	7	5	1	40

TABLE IX. 15-NODE CALCULATION EXAMPLE 3PL SUPPLIER RELATED INFORMATION

Initial	End	3PL Number	Transportation cost	The mean of random time	The variance of random time
s	1	1	20	12	16
s	1	2	18	15	25
s	1	3	24	10	4
s	2	1	18	16	16
s	2	2	17	17	9
s	2	3	19	15	25
s	3	1	19	14	9
s	3	2	18	15	25
s	3	3	20	14	4
1	4	1	10	8	1
1	4	2	11	6	4
1	5	1	10	8	9
1	5	2	12	7	1

Fig. 9 depicts the detail, it is the path diagram corresponding to the solution result of 15-node. Among them, the black pentagram *s* represents the initial node city, and the black pentagram t is the destination node city. The node cities marked with red pentagrams 2, 4, and 7 are transfer node cities that do not meet the constraint of the number of infected people, that is, transfer node city 2, node city 4, and node city 7 are in high-risk areas, so the path cannot pass through node city 2, node city 4, and node city 7. The remaining transfer node cities that meet the constraint of the number of infected people. The path marked with the blue thick line is the distribution path corresponding to the minimum CVaR that satisfies the cost and the infection risk constraints obtained, and the detailed data is depictd in Table X.

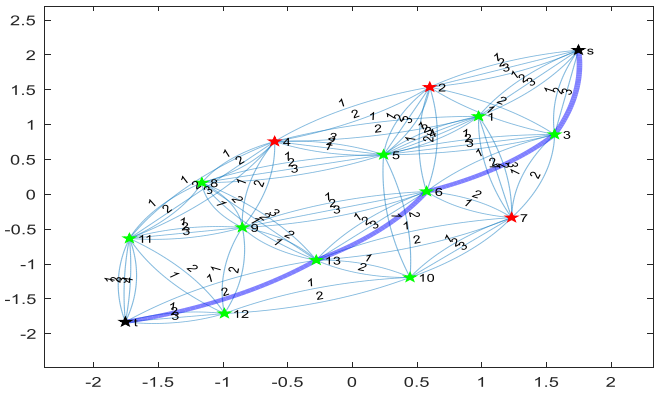

Fig. 9. 15-node solution path diagram.

The solution of the CVaR model of the 15-node problem are given by Table X under different confidence levels, when the cost constraint is C_0=115 and the maximum acceptable the infection risk given by the customer is 1.6×10^{-4}。

TABLE X. SOLUTION OF 15-NODE PROBLEM WHEN T_0=70、C_0=115、F_0=1.6×10^{-4}

β	eposide	CVaR	Best path	F	Best rate	Time
0.9	200	12.1589	R={v_s,3,v_3, 2,v_6,3,v_{13},2,v_t}	1.28×10^{-4}	0.98	0.9s
0.95	200	14.2909	R={v_s,3,v_3,2, v_6,3,v_{13},2,v_t}	1.28×10^{-4}	0.98	0.9s
0.99	200	18.4651	R={v_s,3,v_3, 2,v_6,3,v_{13},2,v_t}	1.28×10^{-4}	0.98	0.9s

It can be known from the data in Table X that when the confidence level is 0.9, the cost constraint C_0=115, and the infection risk constraint F_0=1.6×10^{-4}, the obtained optimal CVaR value is 12.1589, indicating that the corresponding average delay risk of the distribution task is 12.1589. The corresponding distribution cost is 115, and the infection risk is 1.28×10^{-4}. The corresponding best distribution path is $R=\{v_s,3,v_3,2,v_6,3,v_{13},2,v_t\}$, indicating that when transporting from the source node city s to the destination node city t, the selected transfer node cities are 3, 6, and 13 respectively, and the number of the 3PL supplier selected between each two node cities is 3, 2, 3, 2. When the confidence level is 0.95, the cost constraint C_0=115, and the infection risk constraint F_0=1.6×10^{-4}, the obtained optimal CVaR value is 14.2909, indicating that the corresponding average delay risk of the distribution task is 14.2909. The corresponding distribution cost is 115, and the infection risk is 1.28×10^{-4}. The corresponding best distribution path is $R=\{v_s,3,v_3,2,v_6,3,v_{13},2,v_t\}$, indicating that when transporting from the source node city s to the destination node city t, the selected transfer node cities are 3, 6, and 13 respectively, and the number of the 3PL supplier selected between each two node cities is 3, 2. When the confidence level is 0.99, the cost constraint C_0=115, and the infection risk constraint F_0=1.6×10^{-4}, the obtained optimal CVaR value is 18.4651, indicating that the corresponding average delay risk of the distribution task is 18.4651. The corresponding distribution cost is 115, and the infection risk is 1.28×10^{-4} The corresponding best distribution path is $R=\{v_s,3,v_3,2,v_6,3,v_{13},2,v_t\}$, indicating that when transporting from the source node city s to the destination node city t, the selected transfer node cities are 3, 6, and 13 respectively, and the number of the 3PL supplier selected between each two node cities is 3, 2.

The above data indicates that when the cost constraint and the infection risk constraint remain unchanged, as the confidence level increases, the corresponding optimal distribution path will not change, so the infection risk faced will not change either. However, the average delay risk faced by customers will be higher. Therefore, by using this model, 4PL suppliers can combine the customers' aversion to risk and

consider the impact of the infection risk on the distribution plan, so that the distribution path does not pass through high-risk areas, and at the same time, provide customers with the distribution path with the smallest average delay risk that meets the customer's infection risk and cost requirements under the given confidence level. It not only reduces the risk of virus transmission, but also provides customers with a green and efficient delivery solution with the lowest delay risk under a specific confidence level, helping the logistics industry move towards a safer and more sustainable future.

The confidence level selected by clients is significantly influenced by their risk preferences. A higher confidence level may be chosen by clients who are averse to delay risks in order to enhance security, whereas clients with a propensity for taking risks may opt for a lower confidence level to increase the diversity of viable routes. Consequently, our plan is meticulously aligned with the clients' risk preferences, enabling the formulation of bespoke control schemes. Utilizing this plan, 4PL providers can fully take into account the clients' aversion to delay risks, devising distribution routes that circumvent high-risk zones and achieve minimization of the mean delay risk under the specified confidence level. This approach not only mitigates the infection risk but also delivers a green and efficient distribution solution with the minimal delay risk at a particular confidence level, thereby aiding the logistics industry in forging a safer, more efficient, and sustainable distribution system.

C. Algorithm Comparison

In this paper, two distinct algorithms were utilized to tackle the 4PLROP: the Genetic algorithm embedded with the Dijkstra algorithm and the improved Q-learning algorithm. With the aim of assessing the efficacy of these algorithms across varying problem scales, three case studies of differing magnitudes were selected for analysis, encompassing 7 nodes, 15 nodes, and 30 nodes respectively. A comparative examination of the solution outcomes derived from these algorithms on the aforementioned case studies facilitates a profound comprehension of their divergent performances in terms of solution efficiency, solution quality, and stability. This, in turn, furnishes a more efficacious basis for algorithm selection in addressing real-world logistics routing optimization issues. Subsequently, in an endeavor to further scrutinize whether the improved Q-learning algorithm exhibits significant performance enhancements over the traditional Q-learning algorithm, a comparative study was undertaken between the improved Q-learning algorithm and the traditional Q-learning algorithm.

The Genetic algorithm embedded with the Dijkstra algorithm and the improved Q-learning algorithm are used to solve three examples of different scales. The comparison data are demonstrated in Table XI. It can be known from the data in Table XI that when solving the small scale problem of 7-node, both the improved Q-learning algorithm and the Genetic algorithm embedded with the Dijkstra algorithm can find the optimal solution, but the latter shows an inferior solving speed. With the increase of the solving scale, the improved Q-learning algorithm presents a higher solving speed and solving quality. Although the Genetic algorithm embedded with the Dijkstra algorithm can find the optimal solution, the number of iterations and time increase significantly. The main reason is that in this algorithm, a simple graph is first generated, and the Dijkstra algorithm is used to find the optimal path on the simple graph, and then the Genetic algorithm is used for optimization. This leads to the possibility that different simple graphs may find the same path, thereby delaying the optimization convergence process and rapidly increasing the solving time.

TABLE XI. COMPARISON OF RESULTS OF DIFFERENT ALGORITHMS

Node	Algorithm	CVaR	Best path	Best rate	Time
7	Improved Q-learning	22.0456	R={v_s,2,v_2,2, v_3,1,v_t}	1	0.9s
	Embedded Dijkstra's Genetic Algorithm	22.0456	R={v_s,2,v_2,2, v_3,1,v_t}	0.95	19.4s
15	Improved Q-learning	3.7728	R={v_s,1,v_2,2, v_6,3,v_{13},2,v_t}	0.98	1s
	Embedded Dijkstra's Genetic Algorithm	3.7728	R={v_s,1,v_2,2, v_6,3,v_{13},2,v_t}	0.94	24.5s
30	Improved Q-learning	13.641	R={v_s,1,v_4,2,v_8,1, v_{12},1,v_{15},2,v_{18},4, v_{21},1,v_{25},1,v_t}	0.95	1.5s
	Embedded Dijkstra's Genetic Algorithm	13.641	R={v_s,1,v_4,2,v_8,1, v_{12},1,v_{15},2,v_{18},4, v_{21},1,v_{25},1,v_t}	0.9	28.5s

Fig. 10-12 provide a visual representation, they respectively represent the comparison curves of the traditional Q-learning algorithm and the improved Q-learning algorithm when solving 7-node, 15-node and 30-node. In the table, iQlearning refers to improved Q-learning algorithm.

Fig. 10-12 offer a graphical summary of the results, the red curve represents the solution curve of the traditional Q-learning algorithm, and the exploration strategy adopted is the ε-greedy strategy, where the value of ε is 0.5; the green curve represents the solution curve of the improved Q-learning algorithm described in this paper, and the strategy adopted is to randomly select actions with a reward value greater than -1 in the current state.

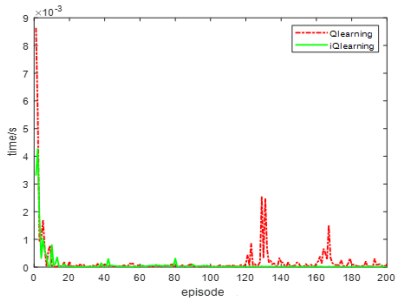

Fig. 10. Comparison of two algorithms at 7-node.

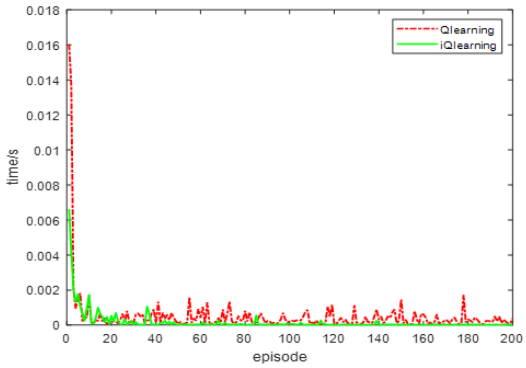

Fig. 11. Comparison of two algorithms at 15-node.

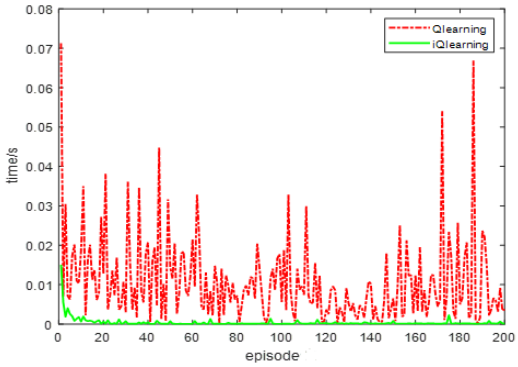

Fig. 12. Comparison of two algorithms at 30-node.

When solving the 7-node problem, both the improved Q-learning algorithm and the traditional Q-learning algorithm can converge relatively quickly, but the former has higher stability and a relatively faster convergence speed. As the problem scale increases, the improved Q-learning algorithm shows higher stability and solution speed. Through the performance of the traditional Q-learning algorithm and the improved Q-learning algorithm in different examples in the above text, it can be seen that the improved algorithm has a faster exploration speed, stronger stability, and a faster convergence speed, which has strong practical significance in solving the 4PLROP.

In summary, this paper has drawn the following conclusions through a comparative analysis of the solution performance of the Genetic algorithm embedded with the Dijkstra algorithm and the improved Q-learning algorithm on case studies of varying scales, as well as the performance disparities between the improved Q-learning algorithm and the traditional Q-learning algorithm: The improved Q-learning algorithm has demonstrated significant performance advantages in addressing the 4PLROP, whether it be in small-scale or large-scale issues, including a faster solution speed, a higher solution quality, and a stronger stability. This indicates that the improved Q-learning algorithm is an effective and practical solution method, capable of providing robust support for path optimization in actual logistics distribution. In the future, we will continue to conduct in-depth research and optimization of this algorithm to further enhance its adaptability and solution efficiency in complex logistics environments, thereby offering a more comprehensive solution for the resolution of 4PLROP.

VI. Conclusions

In the context of a severe pandemic environment, the stability and efficiency of logistics services are crucial for ensuring the continuity of societal operations and the well-being of the populace. Particularly against the backdrop of the dual carbon goals (Carbon Peak and Carbon Neutrality), it is imperative not only to meet the demands of minimizing the risk of delivery delays for clients but also to give due consideration to the prevention and control of infection risks, as well as the imperatives of energy conservation and emission reduction. Together, these efforts weave a logistics network that is green, secure, and efficient, contributing to the sustainable development of the planet.

The present article introduces the CVaR measure and constructs a novel mathematical model. This model not only incorporates distribution costs and infection risks as significant constraints but also aims to minimize the CVaR as the optimization target. Through this model, it ensures that distribution plans can meet the requirements of cost effectiveness and risk control under the complex and variable epidemic environment. To satisfy the demands of this model, the reward and punishment mechanisms in the Q-learning algorithm have been redesigned to more accurately reflect the various risks and cost factors in the actual distribution process. By solving different cases, the lowest CVaR distribution routes that meet the requirements of cost and customer infection risks are obtained. Customers can obtain multiple schemes according to their risk preferences and take corresponding measures. This article provides scientific decision making basis and efficient and safe distribution plans for the 4PL, promoting the logistics industry to move towards a low-carbon, environmentally friendly, and sustainable direction.

Meanwhile, the probability distribution of the stochastic distribution time and transit time is known in this study. When these parameters are unknown, our research is intended to be extended to a robust 4PLROP considering infection risk and delay risk. We anticipate that the improved Q-learning algorithm and the Genetic algorithm embedded with the Dijkstra algorithm will continue to shine brightly. Their potential in 4PLROP is boundless. Moreover, we envision a future where they are seamlessly integrated into various other crucial aspects of the logistics and transportation ecosystem.

Acknowledgment

This research was funded by the Natural Science Foundation of Liaoning Province, grant number 2024-BS-227; the Foundation of the Educational Department of Liaoning Province, grant number No. JYTQN2023345; the Social Science Foundation of Liaoning Province, grant number L24CGL031.

References

[1] M. Huang, L. W. Dong, H. B. Kuang, Z. Z. Jiang, L. H. Lee, X. W. Wang, "Supply chain network design considering customer psychological behavior—a 4PL perspective," Comput. Ind. Eng, pp. 159, 2021.

[2] M. Q. Yin, M. Huang, X. H. Qian, D. Z. Wang, X. W. Wang, L. H. Lee, "Fourth-party logistics network design with service time constraint under stochastic demand," J. Intell. Manuf, vol. 34, pp. 1203–1227, 2023.

[3] D. G. Mogale, D. Xian, V. Sanchez Rodrigues, "Managing logistics risks in pharmaceutical supply chain: a 4PL perspective," Prod. Plan. Control, pp. 1–16, 2024.

[4] K. Kang, R. Y. Zhong, S. X. Xu, "Auction-based cloud service allocation and sharing for logistics product service system," J. Clean. Prod, vol. 278, 2021.

[5] F. Q. Lu, H. L. Bi, W. J. Feng, Y. L. Hu, S. X. Wang, X. Zhang, "A Two-Stage Auction Mechanism for 3PL Supplier Selection under Risk Aversion," Sustainability, vol. 13, pp. 9745, 2021.

[6] M. B. Çağlar Kalkan, K. Aydın, "The role of 4PL provider as a mediation and supply chain agility," Mod. Supply Chain Res. & Appl, vol. 2, pp. 99–111, 2020.

[7] D. Werf, "Information Technology and Data Use in 1PL-4PL Logistic Companies," 2021.

[8] Y. Tao, E. P. Chew, L. H. Lee, Y. R. Shi, Y. Tao, E.P. Chew, L.H. Lee, Y.R. Shi, "A column generation approach for the route planning problem in fourth party logistics," J. Oper. Res. Soc, vol. 68, pp. 165–181, 2017.

[9] H. Pavlić Skender, P. A. Mirković,I. Prudky, "The role of the 4PL model in a contemporary supply chain," Pomorstvo, vol. 31, no. 2, pp. 96–101, 2017.

[10] M. Huang, L. Ren, L. Hay. Lee, X. W. Wang, "4PL routing optimization under emergency conditions,"Knowl. - Based Syst, vol. 89, pp. 126–133, 2015.

[11] M. Huang, L. Ren, L. H. Lee, X. W. Wang, H. B. Kuang, H. B. Shi, "Model and algorithm for 4PLRP with uncertain delivery time," Inf. Sci, vol. 330, pp. 211–225, 2016.

[12] R. R. Ren, Y. F. Zhao, F. Q. Lu, M. Feng, "Research on 4PL Routing Problem Considering Customer Risk Preference," Math. Pract. Theory, vol. 53, no. 1, pp. 163–174, 2023.

[13] W. Hong, Z. l. Xu, W. Liu, L. H. Wu, X. J. Pu, "Queuing theory-based optimization research on the multi-objective transportation problem of fourth party logistics," Proc. Inst. Mech. Eng. B, vol. 235, no. 8, pp. 1327–1337, 2021.

[14] S. H. Yang, J. Zhu, Q, Wang, M. Huan, "Routing Problem with Stochastic Delay Time Under Forth Party Logistics," Proc. 32nd Chinese Control Decis. Conf, vol. 5, no. 6, 2020.

[15] W. Y. Cai, X. F. Wang, X. H. Qian, M. Q. Yin and Y. X. Li, "A route problem with customers' preferences for a fourth party logistics provider," CCDC. Kunming. China, 2021, pp. 4185–4189, 2021.

[16] A. Tatarczak, "A Framework to Support Coalition Formation in the Fourth Party Logistics Supply Chain Coalition," Acta Univ. Lodz. Folia Oecon, vol. 5, no. 338, pp. 192–212, 2018.

[17] S. H. Zhou, Q. Wang, Y. B. Sun, M. T. Yang, D. X. Li, M. Huang, "A Combinatorial Optimization Model for Customer Routing Problem Considering Cost Discount," CCDC. Hefei. China, pp. 2210–2214, 2020.

[18] M. Huang, L. Ren, L. H. Lee, X. W. Wang, "4PL routing optimization under emergency conditions," Knowl. - Based Syst, vol. 89, pp. 126–133, 2015.

[19] M. Huang, L. Ren, L. H. Lee, X. W. Wang, "Model and algorithm for 4PLRP with uncertain delivery time," Inf. Sci. vol. 330, pp. 211–225, 2016.

[20] F. Q. Lu, H. L. Bi, L. Huang , W. Bo, "Improved genetic algorithm based delivery time control for Fourth Party Logistics," 2017 13th IEEE CASE. Xi'an. China, pp. 390–393, 2017.

[21] F. Q. Lu, W. J. Feng, M. Y. Gao, H. L. Bi, S. X. Wang, "The Fourth-Party Logistics Routing Problem Using Ant Colony System-Improved Grey Wolf Optimization," J. Adv. Transp, vol. 1, October 2020.

[22] F. Q. Lu, W. D. Chen, W. J. Feng, H. l. Bi, "4PL routing problem using hybrid beetle swarm optimization," SOFT COMPUT, vol. 27, pp. 17011–17024, 2023.

[23] R. Liang, F. Qing, S. Yuan, L. Ao, "Fourth Party Logistics Routing Problem with Time Window in Uncertain Environments," Inf. Control, vol. 47, no. 5, pp. 583–588, 2018.

[24] X. Y. Gao, X. Gao,Y. Liu. "Fourth Party Logistics Routing Problem Under Uncertain Time and Random Demand[J]. Journal of Uncertain Systems," 2024.

[25] L. Ren, Z. R. Zhou, Y. P. Fu, A. Liu, Y. F. Ma, "Integrated optimization of logistics routing problem considering chance preference," Mod. SCRC Appl, 2024.

[26] L. S. Deng, M. Huang, D. Z. Wang, M. Q. Yin. "Multi-point to multi-point multi-task fourth party logistics routing problem considering tardiness risk," IEEE CASE. C, pp. 1350–1355, 2017.

[27] G. H. Bo, M. Huang, "Model and Solution of Routing Optimization Problem in the Fourth Party Logistics with Tardiness Risk,"ComplexSyst. Complex. Sci, vol. 15, no. 03, pp. 66–74, 2018.

[28] W. Wang, M. Huang, X. W. Wang, "The Optimization Model of 4PL Routing Problem for Risk-averse Customer," CCDC. Hefei. China, pp. 2215–2219, 2020.

[29] X. Liu, G. H. Bo, "Q-Learning Algorithm for Fourth Party Logistics Route Optimization Considering Tardiness Risk," Proceedings of the 2022 International Conference on Cyber-Physical Social Intelligence, 2022.

[30] J. Tu, L. D. Wan, Z. J. Sun, "Safety Improvement of Sustainable Coal Transportation in Mines: A Contract Design Perspective," Sustainability, vol. 15, no. 3, pp. 2085–2095, 2023.