



# Multimodal transportation routing optimization based on multi-objective $Q$ -learning under time uncertainty

Tie Zhang<sup>1</sup> · Jia Cheng<sup>1</sup> · Yanbiao Zou<sup>1</sup>

Received: 28 June 2023 / Accepted: 2 December 2023 / Published online: 16 January 2024  
© The Author(s) 2024

## Abstract

Multimodal transportation is a modern way of cargo transportation. With the increasing demand for cargo transportation, higher requirements are being placed on multimodal transportation multi-objective routing optimization. In multimodal transportation multi-objective routing optimization, in response to the limitations of classical algorithms in solving large-scale problems with multiple nodes and modes of transport, the limitations of directed transportation networks in the application, and the uncertainty of transport time, this paper proposes an optimization framework based on multi-objective weighted sum  $Q$ -learning, combined with the proposed undirected multiple-node network, and characterizes the uncertainty of time with a positively skewed distribution. The undirected multiple-node transportation network can better simulate cargo transportation and characterize transfer information, facilitate the modification of origin and destination, and avoid suboptimal solutions due to the manual setting of wrong route directions. The network is combined with weighted sum  $Q$ -learning to solve multimodal transportation multi-objective routing optimization problems faster and better. When modeling the uncertainty of transport time, a positively skewed distribution is used. The three objectives of transport cost, carbon emission cost, and transport time were studied and compared with PSO, GA, AFO, NSGA-II, and MOPSO. The experimental results show that compared with PSO, GA, and AFO using a directed transportation network, the proposed method has a significant improvement in optimization results and running time, and the running time is shortened by 26 times. The proposed method can better solve the boundary of the Pareto front and dominate the partial solutions of NSGA-II and MOPSO. The effect of time uncertainty on the performance of the algorithm is more significant in transport orders with high time weight. With the increase in uncertainty, the reliability of the route decreases. The effectiveness of the proposed method is verified.

**Keywords** Multimodal transportation · Multi-objective  $Q$ -learning · Weighted sum · Undirected multiple-node transportation network · Time uncertainty

## Introduction

Multimodal transportation [1, 2] is a modern way of transporting cargo by synergizing two or more modes of transport. Compared with traditional unimodal transportation, multimodal transportation can make full use of the advantages of various modes of transport, thus reducing logistics costs

and improving logistics efficiency [3]. In addition, multimodal transportation can reduce pollution and has better environmental benefits [4]. However, in practical application, the advantages of multimodal transportation can only be better utilized if suitable transport modes and routes are selected. Multimodal transportation involves multiple objectives. To find the optimal solution, accurate modeling of the transportation network is necessary. Moreover, there are uncertainties in the actual transportation process. Therefore, this paper studied multimodal transportation multi-objective routing optimization, the establishment of a transportation network, and the uncertainty of transport time.

In multimodal transportation, there are multiple participants involved, such as customers and carriers. Each participant has different goals, such as the shortest transport time, the lowest transport costs, and the least carbon

---

✉ Tie Zhang  
merobot@scut.edu.cn

Jia Cheng  
cchengjia@foxmail.com

Yanbiao Zou  
ybzou@scut.edu.cn

<sup>1</sup> School of Mechanical and Automotive Engineering, South China University of Technology, Guangzhou 510641, China

emissions. With multiple delivery time options, for example, the same day, tomorrow, or the day after tomorrow, customers expect cargo to be delivered on time within their choice. Carriers expect to reduce transportation costs as much as possible while meeting the customers' choices. In addition, carbon emissions taxes have become factors to be considered in light of global climate issues and increased awareness of environmental protection [5]. There are conflicts and contradictions among various objectives, so the multimodal transportation problem is often established as a multi-objective optimization problem, aiming to find an optimal solution that balances various objectives and achieves the maximum comprehensive benefits and sustainable development. For example, Resat et al. [6] considered the two objectives of time and cost to establish the multimodal transportation problem in the Marmara region of Turkey as a multi-objective optimization problem. Zheng et al. [7] considered the three objectives of time, cost, and carbon emission to establish the multi-objective optimization problem of multimodal transportation.

In multimodal transportation, there are many widely used optimization algorithms such as mathematical programming [8], genetic algorithms [9], particle swarm algorithms [10], and ant colony algorithms [11]. Mathematical programming algorithms are fast. Heuristic algorithms have the advantages of good optimization results and wide applicability. However, there are limitations in the application of widely used algorithms. Heuristic algorithms require carefully adjusting parameters to avoid falling into local optima. When the scale of a multi-objective optimization problem is too large, mathematical programming methods may not be able to solve it, and the performance of the heuristic algorithm may reduce.

In recent years, reinforcement learning has been successfully applied to route planning [12, 13], transport scheduling [14], and loading optimization [15]. Reinforcement learning is an algorithm that continuously interacts with the environment to learn the best action strategy and can be used to solve sequential decision problems. In the field of multimodal transportation route optimization, there is little research on the application of reinforcement learning. However, with the increase of multimodal transportation problems scale and search space complexity, widely used algorithms have difficulty in solving them, and reinforcement learning have more advantages in such situations. In addition, the adaptive nature of reinforcement learning, which allows for autonomous optimization of decisions based on changes in the environment, is a good fit for the uncertainty that exists in multimodal transportation problems. In practical applications, the convergence of an algorithm is crucial to its successful application, and in the field of reinforcement learning, many studies have demonstrated that the  $Q$ -learning algorithm has good convergence performance

[16–19]. Therefore, in this paper, the  $Q$ -learning algorithm is used to solve the multimodal route optimization problem.

In the field of multi-objective optimization (MOO), there are two common strategies for converting a multi-objective to a single-objective, and for solving a Pareto frontier based on a dominance relation. Like the MOO problem, multi-objective reinforcement learning (MORL) [20, 21], can be classified into single-strategy multi-objective reinforcement learning and multi-strategy multi-objective reinforcement learning based on the strategies. The former converts multi-objective reinforcement learning into single-objective reinforcement learning and uses single-objective reinforcement learning algorithms to find the best solution. For single-strategy multi-objective  $Q$ -learning, the weighted sum approach [22], the W-learning approach, the analytic hierarchy process approach [23], and the ranking approach [24], etc. can be used. The weighted sum approach ignores the different units and ranges between different objective functions and sets a weight vector based on the preference between multiple objectives. Based on the weight vector, the individual reward functions are weighted and summed directly to obtain the final scalar reward function, thus converting the multi-objective optimization problem into a single-objective optimization problem. By setting multiple sets of weights, multiple sets of solutions can be found. As the implementation process can easily adjust the weights according to the importance of the objectives, the weighted sum approach is now widely used. For example, Zeng et al. [25] used the weighted sum to combine three objectives and applied it to a reinforcement learning algorithm. Ngai et al. [22] used weighted sum to combine seven objectives and applied it to a reinforcement learning algorithm for implementing vehicle overtaking. Therefore, in this paper, the weighted sum  $Q$ -learning is used to solve the multimodal multi-objective route optimization problem.

To ensure that the algorithm can find the optimal solution, an accurate model of the multimodal transportation network is necessary. In common optimization algorithms, multimodal networks are modeled as containing origins, destinations, and multiple intermediate nodes, and usually with only one direction of transport, i.e., a directed transportation network. Zhang et al. [26] have established a directed multimodal transportation network from Nanjing to Harerbin, containing 13 intermediate nodes. Sun [27] has established a directed multimodal transportation network containing 12 intermediate nodes. However, there are limitations to the directed modeling approach. As the transport direction is directional, if the manual setting of route directions is wrong, the suboptimal solution may be caused. In practical logistics applications, the transportation of cargo is usually undirected. Also, the directed modeling approach needs to modify or re-establish the network when the origin and destination of the cargo change. In addition, the directed modeling approach

mostly uses the multiple-edge approach, which cannot effectively represent transit information.

To solve the above problems, a more accurate, realistic, and convenient undirected multiple-node multimodal transportation network model is established, which has the following advantages: (1) there is no need to set the transport direction of the route manually, only need to represent the path all over the network. Therefore, it can avoid suboptimal solutions due to the wrong route direction set manually. In addition, when the origin and destination of transported cargo change, our network can still work, and there is no need to modify or re-establish the network. (2) The mode of transport between nodes in our network is undirected, which makes our network more realistic and flexible. (3) Using the multiple-node approach, the network is established as a multiple-node network. That is, each mode of transport of each node corresponds to a sub-node, which can better represent the transfer information. Using the undirected multiple-node modeling approach allows for a more realistic representation of transportation in a multimodal transportation network. The undirected transport direction provides more options and possibilities for the algorithm when finding the optimal solution.

During the process of establishing the model, information such as transport costs and transport time of the routes need to be taken into account. Multimodal transportation is a complex system. Transportation may be affected by weather, traffic, and other factors, which may lead to changes in transport times, transport costs, and so on. Therefore, uncertainty is one of the unavoidable factors in multimodal transportation problems. If the model does not take into account the uncertainty factor, the obtained optimal route will lose its meaning in the actual transportation. Among the existing studies, stochastic theory [28–30], fuzzy theory [27], or robust optimization [31] are effective methods to solve the uncertainty problem in multimodal transportation. Baykasoğlu et al. [32] established the demand for import and export freight as a random number. Haddadsisakht et al. [33] considered the randomness of the carbon tax. Zhang [26] treated the stochastic transport time as a set of random numbers following a normal distribution. Demir [29] considered the uncertainty of transport times and built them as random numbers and applied them to a real network model. Experimental results showed the advantage of randomness in generating robust transport solutions, outperforming deterministic models. In cargo transportation, the length of transport time will affect customer satisfaction and hence the carrier's interest. Therefore, the network built in this paper takes into account the uncertainty of the transport time, which is established as a random number following a positively skewed distribution according to the actual situation.

In summary, to address the limitations of widely used optimization algorithms, the limitations of directed network

modeling, and the uncertainty in transportation networks in the field of multimodal transportation multi-objective routing optimization. This paper proposes an optimization framework based on a multi-objective  $Q$ -learning algorithm, which uses a weighted sum approach to solve multiple-objective problems, establishes an undirected multiple-node multimodal transportation network, and uses random numbers to represent time uncertainty, providing more options and possibilities for the solution of the algorithm.

## Contribution

1. An undirected multimodal transportation multiple-node network is proposed to solve the problem of suboptimal solutions due to manually set routes in the wrong direction, and the need to modify or re-establish the network when the origin or destination of cargo transportation changes. The transport route of this network is undirected, with each node representing multiple transport modes through multiple sub-nodes. The undirected transport routes can provide more options for algorithmic optimization, while the introduction of multiple-node brings convenience to the representation of transfer information.
2. To characterize the uncertainty of transport time, a multimodal transport model with time uncertainty is constructed based on a positively skewed distribution.
3. To solve the uncertain multimodal transport route optimization problem, combined with the established network, an optimization framework based on multi-objective weighted sum  $Q$ -learning is proposed, applying  $Q$ -learning to complex multimodal and uncertain routing problems, and the convergence is analyzed.

## Problem statement and model

In this section, the design and establishment of the multimodal transportation network model is discussed and the multimodal transportation mathematical model applicable to  $Q$ -learning is defined.

### Problem statement

In multimodal transportation, cargo can be transported from the origin to the destination through multiple modes of transport, which involves combined transport between different modes of transport. In China, waterway transport is developed, but the level of railway–waterway intermodal transport and highway–waterway intermodal transport is low. In terms of railway–waterway intermodal transport, the proportion

is only 2.6%. Moreover, many cities in the multimodal transportation network are non-coastal cities. Therefore, this paper does not consider waterway transport, only considering the highway, railway, and airway three modes of transport.

Due to various factors, there are many uncertainties in transportation, such as demand uncertainty, transport time uncertainty, and transport capacity uncertainty. Increased transport time for trucks due to traffic jams, or delayed trains and planes can lead to uncertain transport time. This can lead to cargo arrival times exceeding customer expectations or requirements, reducing customer satisfaction and affecting the carrier's interests. Therefore, this paper considers the uncertainty of transport time.

The goal of multimodal transportation is to maximize the efficiency of transportation, that is, to spend as little cost and time as possible, to minimize carbon emissions, and to choose an optimal route to get cargo from the origin to the destination. When there are a large number of transport nodes, the global optimization of multiple objectives is a typical NP-HARD problem. At this time, the carrier needs to weigh and choose and set weights for each objective according to its importance.

Therefore, this paper establishes a multimodal transportation network  $G = (V, E, M)$ , where  $V$  is the set of nodes of the network,  $E$  is the set of edges of the network, and  $M$  is the set of transport modes of the network (i.e. highway transport, railway transport, airway transport). And the multimodal transport model under transport time uncertainty is constructed. The objective function is constructed as a weighted sum of multiple sub-objectives. And the multi-objective multimodal transportation routing optimization problem is described as an optimization model with a minimized objective function. In the established network, the optimal transport routes and transport modes are selected according to the objective function.

## Modeling of multimodal transportation network

In the realistic transportation network, there are many transport modes for cargo transportation between each city, i.e., transport nodes. In some papers [26, 34], nodes are not necessarily connected, i.e., if one wants to get from one node to another, one must pass through other nodes. In general, cities can be connected by highway transport, and one can reach another city directly from one city. That is, there can be at least highway transport modes between the nodes of the network.

In some papers, the transport routes of the established transportation network are directed, as shown in Fig. 1. The routes of each city lead to other cities and eventually reach the destination, which facilitates the algorithm to find the optimal route. However, it requires a manual setting of the transport direction of the route, which may lead to suboptimal

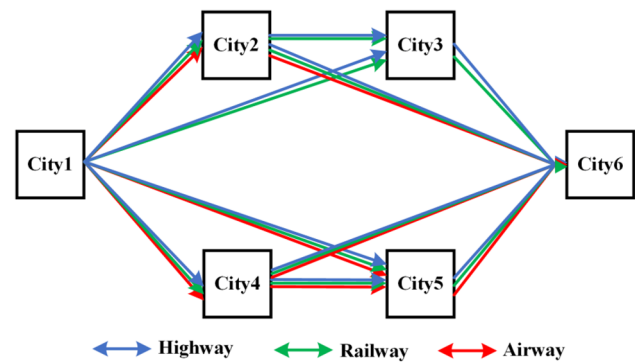


Fig. 1 Schematic diagram of multimodal directed network

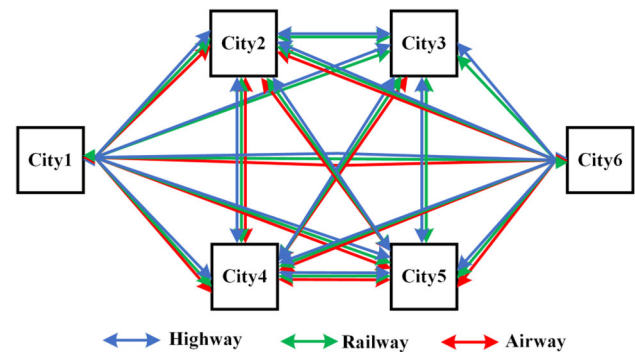


Fig. 2 Schematic diagram of multimodal undirected network

solutions when the setting is wrong. Moreover, this does not match the realistic transportation network, where the transport route is mostly undirected. In addition, when the origin and destination of the transportation are modified, the transportation network is needed to be modified or re-established.

To solve the above problems, and to establish a network that can better reflect the transportation of cargo, this paper establishes a more convenient, realistic, and flexible undirected multimodal transportation network model. The network diagram is shown in Fig. 2, where all nodes are connected and the transport route is undirected, which also facilitates the modification of the origin and destination of cargo.

## Mathematical model

### Problem assumptions

To facilitate the study, the following model assumptions are proposed.

1. The same batch of cargo is indivisible in the transportation process, i.e., it can only be transported as a whole and cannot be divided into two or more parts during the transportation process.



2. The unit transport cost and carbon emission between different nodes are known, the unit transfer time, unit transfer cost, and unit transfer carbon emission between different transport modes are known, and the average speed of different transport modes is known.
3. The random distribution of transport time between different nodes is known and follows the positively skewed distribution.
4. If the cargo arrives at a node and needs to be transferred, the loading and unloading of the cargo will start immediately. The cargo is dispatched immediately upon completion of the transfer.
5. Transfer can only occur at a transport node and the mode of transport for each shipment can only be changed at most once at that node.

### Definition of parameters and variables

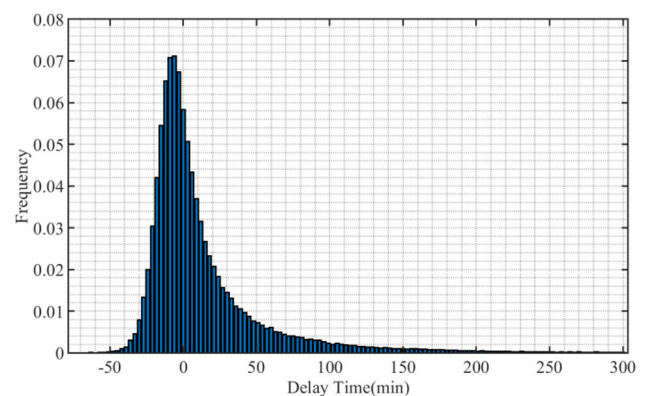
In this paper, the parameters and variables defined in Table 1 are used to construct a mathematical model of the multimodal transport path problem.

### Time uncertainty description

In the realistic transportation network, the transport time of the multimodal transportation network is uncertain due to the influence of traffic jams and weather. In stochastic theory, uncertainty is modeled by appropriate random variables that follow well-known theoretical or empirical distributions. And these variables are independent of each other. Specifically, uncertainty can be simulated by taking random values in the probability distribution that a variable follows. Therefore, the transport time can be considered as a random variable and its uncertainty can be modeled using a probability distribution. Then, when solving the problem, a random value is selected from the distribution, thus incorporating uncertainty into the decision-making process. A histogram is drawn using the American Airlines 2015 flight delay dataset in Fig. 3. The skewness of this data distribution is calculated to be 4.957, i.e., the distribution of the data is right-skewed. Because in practice, transport time is often affected by many random factors, such as traffic conditions, weather conditions, and road conditions. These factors may cause transport time to be extended, thus making the transport time data exhibit a right skew.

**Table 1** Parameters and variables

Sets	$V$	The set of nodes of the network
	$E$	The set of edges of the network, $E = \{E_{ij}^m   i, j \in V, m \in M\}$
Parameters	$M$	The transport mode set of the network
	$C_1$	Cost of transportation (¥)
	$C_2$	Total carbon cost of transportation (¥)
	$C_3$	Total transport time (h)
	$Q$	Total weight of cargo (t)
	$c_{ij}^m$	Unit cost of transporting cargo by transport mode $m$ from node $i$ to node $j$ (¥/t km)
	$c_i^{mn}$	Unit cost of transferring cargo from transport mode $m$ to transport mode $n$ at node $i$ (¥/t)
	$t_{ij}^m$	Unit time for cargo to be transported by transport mode $m$ from node $i$ to node $j$ (h/t km)
	$t_i^{mn}$	Unit transfer time used for the transfer of cargo from transport mode $m$ to transport mode $n$ at node $i$ (h/t)
	$e_{ij}^m$	Unit carbon emissions generated by transporting cargo by mode $m$ from node $i$ to node $j$ (kgCO <sub>2</sub> /t km)
	$e_i^{mn}$	Unit carbon emissions from the transfer of cargo from transport mode $m$ to transport mode $n$ at node $i$ (kgCO <sub>2</sub> /t)
	$d_{ij}^m$	Distance the cargo is transported by transport mode $m$ from node $i$ to node $j$ (km)
	$c_e$	Carbon tax per unit of CO <sub>2</sub> (¥/kgCO <sub>2</sub> )
Decision variables	$x_{ij}^m$	The decision variable, which takes the value of 1 when the cargo is transported from node $i$ to node $j$ by mode $m$ and 0 otherwise
	$y_i^{mn}$	The decision variable, which takes the value of 1 when the cargo is transferred from transport mode $m$ to transport mode $n$ at node $i$ and 0 otherwise



**Fig. 3** Flight delay data histogram

For logistics companies, a discrete or continuous distribution of actual transport times can be fitted by using data from historical orders. In this paper, to characterize the uncertainty in transport time, the transport time for each route for each mode of transport is considered a random variable. To better fit the distribution of transport time in practice, it is assumed that the transport time conforms to a right-skewed normal distribution, i.e., a positively skewed distribution and that the random variables are independent of each other.

### Optimization objectives and constraints

In multimodal transportation, cargos need to be transported from origin to destination within a specified time and at the least possible cost. Reducing carbon emissions is also a key objective of modern logistics due to the promotion of a low-carbon environment. In practice, transport modes with low carbon emissions are usually accompanied by low transport costs. Specifically, railway transport has the least carbon emissions and the lowest transport costs. Airway transport produces the most carbon and costs the most. Highway transport is in the middle. In addition, the value of carbon emissions multiplied by the carbon tax gives a very small value for the cost of carbon and will be even smaller when multiplied by the weights, which will make it difficult to influence the result. Moreover, for carriers, a focus on so-called carbon emissions is a focus on carbon taxes and carbon costs, i.e., economic benefits. Therefore, based on the above analysis, carbon emission costs and transport costs are combined into one objective, called total transport costs, converting the three-objective optimization problem into a two-objective optimization problem.

Considering transport cost, carbon emission cost, and transport time, a multimodal transportation model with minimum weighted sum cost is constructed. The objective function is,

$$Z = \omega_1 \cdot (C_1 + C_2) + \omega_2 \cdot C_3 \quad (1)$$

where  $Z$  is comprehensive transportation cost,  $C_1$ ,  $C_2$ , and  $C_3$  are transport cost, carbon emission cost, and transport time, respectively. The total transport cost is the sum of transport cost and carbon emission cost,  $\omega_1$  and  $\omega_2$  are the weights in the objective function, and their sum is 1. The weights can reflect the carrier's preference for total transport cost and transport time. The objective function is to minimize the total weighted cost.

Transport cost  $C_1$  includes the transport cost between nodes  $C_{tp}$  and the transfer cost  $C_{tf}$  at each node.  $C_{tp}$  is the cost incurred in the transport process, which is proportional to the transport distance and the weight of the cargo.  $C_{tf}$  is

the cost incurred due to the change of transport mode, which is proportional to the weight of the cargo.  $C_{tp}$ ,  $C_{tf}$  can be expressed by Eq. (2) and (3), respectively. The transport cost  $C_1$  is the sum of both  $C_{tp}$  and  $C_{tf}$  as shown in Eq. (4).

$$C_{tp} = Q \cdot \sum_{m \in M} \sum_{i \in V} \sum_{j \in V} (c_{ij}^m \cdot d_{ij}^m \cdot x_{ij}^m) \quad (2)$$

$$C_{tf} = Q \cdot \sum_{m \in M} \sum_{n \in M} \sum_{i \in V} (c_i^{mn} \cdot y_i^{mn}) \quad (3)$$

$$C_1 = C_{tp} + C_{tf} \quad (4)$$

The cost of carbon emissions includes carbon emissions from transport between nodes and from the transfer at each node. To implement low-carbon transport and develop a low-carbon economy, many countries have developed carbon tax systems. Therefore, the total carbon emission cost  $C_2$  is the product of the carbon tax  $c_e$  and the total carbon emissions:

$$C_2 = c_e \cdot \left( \sum_{m \in M} \sum_{i \in V} \sum_{j \in V} (e_{ij}^m \cdot d_{ij}^m \cdot x_{ij}^m) + \sum_{m \in M} \sum_{n \in M} \sum_{i \in V} (e_i^{mn} \cdot y_i^{mn}) \right) \quad (5)$$

Transport time  $C_3$  includes the transport time  $T_{tp}$  between nodes and the transfer time  $T_{tf}$  at each node.  $T_{tp}$  can be calculated from the average speed of the modes of transport and the distances between cities, while  $T_{tf}$  depends on the mass of the cargo  $Q$ , and can be obtained by multiplying  $Q$  and the unit time of transfer, as shown in Eqs. (6) and (7), respectively. The transport time is the sum of  $T_{tp}$  and  $T_{tf}$ , as shown in Eq. (8).

$$T_{tp} = \sum_{m \in M} \sum_{i \in V} \sum_{j \in V} (t_{ij}^m \cdot x_{ij}^m) \quad (6)$$

$$T_{tf} = Q \cdot \sum_{m \in M} \sum_{n \in M} \sum_{i \in V} (t_i^{mn} \cdot y_i^{mn}) \quad (7)$$

$$C_3 = T_{tp} + T_{tf} \quad (8)$$

The constraints are:

$$\sum_{m \in M} x_{ij}^m \leq 1, \quad \forall i, j \in V \quad (9)$$

$$\sum_{m \in M} \sum_{n \in M} y_i^{mn} \leq 1, \quad \forall i \in V \quad (10)$$

$$x_{ij}^m \cdot x_{jk}^n = y_{jk}^{mn} \quad \forall i, j, k \in V, \quad \forall m, n \in M \quad (11)$$

$$x_{ij}^m \in \{0, 1\} \quad \forall i, j \in V \quad \forall m \in M \quad (12)$$

$$y_i^{mn} \in \{0, 1\} \quad \forall m, n \in M \quad \forall i \in V \quad (13)$$

Equation (9) guarantees that at most one transport mode can be selected between node  $i$  and node  $j$ . Equation (10) constrains that at most one transfer can occur at a node. Equation (11) requires that if cargo is transferred at a node, the transport mode should be the same before and after that node. Equations (12) and (13) constrain the domain of the definition of the two decision variables according to the definition.

## Multi-objective Q-learning algorithm

Reinforcement learning has been shown to solve challenging problems in some industrial applications with promising results. Therefore, this paper proposes a multi-objective Q-learning-based method to solving multimodal routing optimization problems under uncertainty and establishes a training environment for reinforcement learning intelligence, including state space, decision space, and reward functions. If the network converges, the intelligence reaches the destination state with the sequential decision with the highest reward value.

## The multimodal environment

The reinforcement learning algorithm learns by executing many episodes. In each episode, the agent interacts with the environment by exploring and exploiting it to find a route between the origin and the destination. The environment stores information about each transport node and transport mode and can perform actions that will lead to rewards and new states. For this purpose, the environment has three main functions.

The initialization of uncertainty. Using the positively skewed distribution mentioned in “[Time uncertainty description](#)” section, random numbers are generated randomly to initialize the transport time for each route of the three transport modes (Algorithm 1, lines 2).

Environment initialization. At the beginning of each episode, i.e., each time the agent reaches the destination, the environment is initialized and the cargo is repositioned randomly to one of the states of the original transport node (Algorithm 1, lines 4).

Interaction with the agent. To train the Q-learning model, i.e., the Q-table, there is a constant interaction between the agent and the environment. When the agent is in a state, the agent chooses an action, and the environment calculates the reward generated by that action and moves the agent to the next state. With this interaction, the agent reaches the destination, i.e., the cargo reaches its destination, and ends the episode (Algorithm 1, lines 5–17).

## Establishment of the multiple-node approach

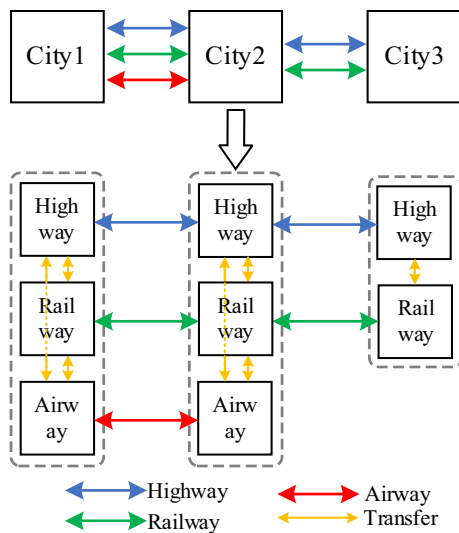
To interact with agents and simulate realistic transportation networks, it is necessary to build multimodal transportation network models, and the construction of a graphical network structure is a very effective way to do this. A graphical network structure consists of nodes and edges, where the nodes are the transport nodes, and the edges are the transport modes and related information between the transport nodes. There are two approaches to modeling multimodal networks, the multiple-node approach and the multiple-edge approach.

In the multi-edge approach, as shown in Fig. 2, a node represents a transport node. When there are multiple modes of transport between two transport nodes, there are multiple edges between two nodes in the network structure, therefore, it is called the multi-edge approach. In practice, the transfer cost is considered. In the multi-edge approach, there may be multiple transport modes to reach a node, and there may also be multiple transport modes from this node to another node. In this case, the transfer cost involves two modes of transport before and after the node, that is, the transfer cost is determined by two actions before and after. This will lead to a problem: The agent chooses the same action in the same state, but ends up with a different reward, because the previous action may be different. This is not feasible in reinforcement learning and does not conform to the Bellman equation.

In response to the above problem, this paper transforms the multiple-edge approach into the multiple-node approach. In the multi-node approach, if a transport node has multiple transport modes, multiple sub-nodes are used to represent that transport node. Each sub-node represents one transport mode of that transport node, indicating reaching the transport node by that transport mode.

In this network, there is at most one edge between sub-nodes of different nodes, representing the transport modes that can be used between two nodes. The edge may describe variables such as the transport cost or transport time of the corresponding transport mode. With this approach, transfer time and transfer cost are easily expressed and understood. At one sub-node, transfer cost or transfer time needs to be considered if the action selected by the agent represents a different mode of transport from that represented by the sub-node, and not otherwise. It facilitates the description of the information related to transfer in reinforcement learning. Therefore, a multimodal transportation network model is developed in this paper using a multiple-node approach.

The handling of multiple nodes is as follows: each transport node is expanded into multiple sub-nodes depending on the number of transport modes it has. For example, the first three transport nodes in Fig. 2 are expanded as shown in Fig. 4, where each sub-node represents one transport mode.



**Fig. 4** Schematic diagram of multiple-node approach

City1 and City2 have three transport modes, so they are expanded into three sub-nodes. City3 has only two transport modes, so it is expanded into only two sub-nodes. There is also a transport relationship between the sub-nodes of the same node, i.e., the transfer of cargo, but the agent does not select the sub-nodes of the same node as the next action, i.e., the sub-nodes of the nodes are in the same level.

### Multi-objective Q-learning

In reinforcement learning, multimodal routing optimization is regarded as a Markov decision process (MDP) [35]. MDP includes states, actions, and rewards, making decision i.e., choosing a certain action in a state, obtaining the corresponding reward, and entering the next state.

According to the multiple-node method of “[Establishment of the multiple-node approach](#)” section, the state,  $s$ , consists of two parts, the first part represents the serial number of the transport node, and the second part represents the transport mode, i.e.,  $s = [num, way]$ , with different transport modes representing different sub-nodes. For example, if the transport node is 1 and the transport mode is the railway, the corresponding state,  $s$ , is  $[1, railway]$ , which represents the arrival of transport node 1 through railway transport.

For the set of actions,  $A$ , which are the edges of a state connected to other states, the number of edges represents the number of actions available in the set. The choice of action represents the choice of a transport mode to reach a node.

For the reward function  $R(s, a)$ , the reward function is related to the current state the agent is in and to the action chosen, independent of the next state. According to the objective function of “[Optimization objectives and constraints](#)” section, the reward function consists of the total transport cost and transport time, which is the negative of the weighted sum of the costs associated with each objective resulting from the selection of action  $s$  from the actions of state,  $s$ .

$$R(s, a) = -(\omega_1 \cdot (C_1(s, a) + C_2(s, a)) + \omega_2 \cdot C_3(s, a)) \quad (14)$$

where  $\omega_1$  and  $\omega_2$  are the weights of each objective, and  $C_1(s, a)$  is the cost generated by transportation, which is the sum of transportation cost  $tp(s, a)$  and transfer cost  $tf(s, a)$ , i.e.,  $C_1(s, a) = tp(s, a) + tf(s, a)$ .  $C_2(s, a)$  is the carbon emission cost generated by transportation, which is the sum of the carbon emission cost of transport  $tp_c(s, a)$  and the carbon emission cost of transfer  $tf_c(s, a)$ , i.e.,  $C_2(s, a) = tp_c(s, a) + tf_c(s, a)$ .  $C_3(s, a)$  is the transport time consumed and is the sum of transport time  $tp_t(s, a)$  between nodes and transfer time  $tf_t(s, a)$ , i.e.,  $C_3(s, a) = tp_t(s, a) + tf_t(s, a)$ .

The objective is to maximize the expected cumulative reward of the selected action, which is equal to minimizing the expected cumulative cost of transportation, calculated using the Bellman equation [36].

It is important to note that in this paper, the reward is not normalized. Since normalizing the reward requires subtracting a constant from the reward, this may affect the reward function and is not consistent with the Bellman equation [37]. Therefore, in this paper, the reward is not normalized. However, since the value of the total transport cost is two orders of magnitude larger than the transport time, the weight of the transport time in the reward function is only greater if  $\omega_2$  is two orders of magnitude larger than  $\omega_1$ .

In summary, the pseudo-code for the method proposed in this paper is as follows:



---

```

1: Initialize  $Q(s, a)$  arbitrarily
2: Initialize  $\varepsilon_{\text{start}}, \varepsilon_{\text{end}}, \varepsilon_{\text{decay}}, \alpha, \gamma, \text{weights}[\omega_1, \omega_2]$ 
2: Generate random time for all transportation modes
3: for episode = 1 to E do
4:   Initialize state,  $s, \varepsilon = \varepsilon_{\text{start}}$ 
5:   while state,  $s$ , is not terminal do
6:     Obtaining the action space  $A$  according to the state  $s$ 
7:     With probability  $\varepsilon$  select a random action  $a \in A$ 
8:     Otherwise select  $a = \arg \max_{a' \in A'} Q(s, a')$ 
9:     Update new state  $s'$  with  $a$ 
10:    Calculate reward  $r = R(s, a)$  according to the random time
11:     $Q(s, a) \leftarrow Q(s, a) + \alpha [r + \gamma \max_{a'} Q(s', a') - Q(s, a)]$ 
12:     $s \leftarrow s'$ 
13:    if  $\varepsilon \cdot \varepsilon_{\text{decay}} > \varepsilon_{\text{end}}$  then
14:       $\varepsilon = \varepsilon_{\text{start}} \cdot \varepsilon_{\text{decay}}$ 
15:    else  $\varepsilon = \varepsilon_{\text{end}}$ 
16:    end if
17:  end while
18: end for

```

---

**Algorithm 1.** Q-Learning for Multimodal Transportation Optimization under Uncertainty

## Convergence analysis

There are rigorous academic proofs of the convergence of  $Q$ -learning [16–19]. According to the literature, to ensure convergence when using  $Q$ -learning in our model, our model needs to satisfy the following assumptions:

1. The model has a finite state space  $s$  and action space  $A$ .
2. Any state space  $(s, a)$  can be visited an infinite number of times.
3. The reward function is independent of the state at the next moment.

For assumption (1), the number of transport nodes, i.e., cities, in this paper's multimodal transportation network model is finite, and the transport modes between each node are also finite. Therefore, the state space  $s$  and action space  $a$  of the model in this paper are finite. The assumption is met.

For assumption (2), in the multimodal transportation network model built in this paper, the transport modes between transport nodes are undirect instead of direct, which is not only more consistent with the realistic situation but also makes any state space  $(s, a)$  can be visited an infinite number of times. Furthermore, in the algorithm, no episode restriction is set in this paper, i.e. any state space  $(s, a)$  can be accessed an unlimited number of times when the agent has not gone to the destination. The assumption is met.

For assumption (3), the reward function established in this paper is only related to the agent's state at that moment and the action choice made, and is not related to the next moment state. The assumption is met.

Therefore, the multimodal  $Q$ -learning model developed in this paper satisfies all the above assumptions. This paper builds a multi-objective  $Q$ -learning model, but the paper uses a weighted sum linear scalar function, which is equivalent to converting the MOMDP to the corresponding MDP, and

**Table 2** Transport mode parameters

Mode of transport	Transport cost (¥/t km)	Average speed (km/h)	Carbon emissions (kgCO <sub>2</sub> /t km)
Roadway	0.35	120	0.0212
Railway	0.165	60	0.0043
Airway	0.6	600	0.1922

the existing convergence proofs still apply [37]. Specifically, Eq. (14) is the set reward function  $R(s, a)$ , where  $\omega_1, \omega_2, C_1(s, a), C_2(s, a), C_3(s, a)$  are all finite values so that the final reward function  $R(s, a)$  is also finite and converts the MOMDP to the corresponding MDP. For the multi-objective  $Q$ -learning in this paper, it still obeys the following update:

$$Q(s, a) = Q(s, a) + \alpha [r + \gamma \cdot \max_{a'} Q(s', a') - Q(s, a)]. \quad (15)$$

Therefore, it is guaranteed that the method proposed in this paper will converge to the best solution.

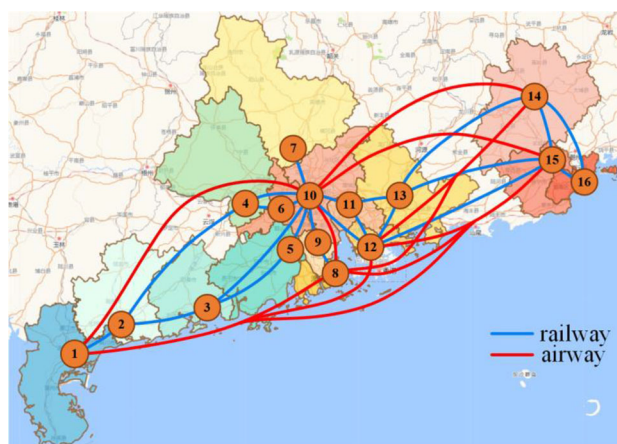
## Case study

In this section, based on the proposed undirected network, two transportation networks are created and the validity is verified. Multi-objective  $Q$ -learning and time uncertainty are studied through cases.

### Transportation network

By referring to the freight tariff table of freight services and reviewing the literature [26], the unit transport cost, average speed, and unit carbon emission of each transport mode are obtained, as shown in Table 2. Assuming that each node satisfies the requirements of transport mode transfer, by reviewing literature [5, 7], the unit transfer cost, time, and carbon emission per unit transfer between different transport modes are combed and obtained, as shown in Table 3.

To better verify the performance of the proposed algorithm in different transportation networks, two transportation networks are set up. In the first transportation network, 16 cities in Guangdong Province, China, are selected as nodes of the transportation network, and the cities and corresponding node numbers are shown in “The first type of transportation network node number and the corresponding city” section. Of these, five cities support airway transport modes. In the second transportation network, 17 cities in eastern China are selected as nodes in the transportation network. Fifteen of the selected cities are provincial capitals, and all of them have both railway and airway transport modes, so the choice

**Fig. 5** Schematic diagram of railway transport and airway transport

of transport modes at the nodes is greater than in the first network, and the scale of the network is larger.

Generally, cities can all be connected by highway transport, either directly from one city to another or via passing through other cities to another city. Therefore, to take both cases into account and to create a network that better reflects the transport of cargo, the highway transport distances from each city node to all other city nodes are collected, which also increases the choice of each node and increases the size of the network. In the Appendix, Tables 10, 11 and 12 list the transport distances of the three modes of transport between each city node of the two transportation networks, where the road distance data is from Gao De Map, the railway distance is from the China Railway website and the airway transport distance is from the flight miles of Southern Airlines. Figure 5 shows the railway transport routes and the airway transport routes of the first transportation network, which are not shown in the figure due to the excessive number of highway transport routes. Figure 6 shows the railway transport routes of the second transportation network, which are not shown in the figure due to the excessive number of highway and airway transport routes.

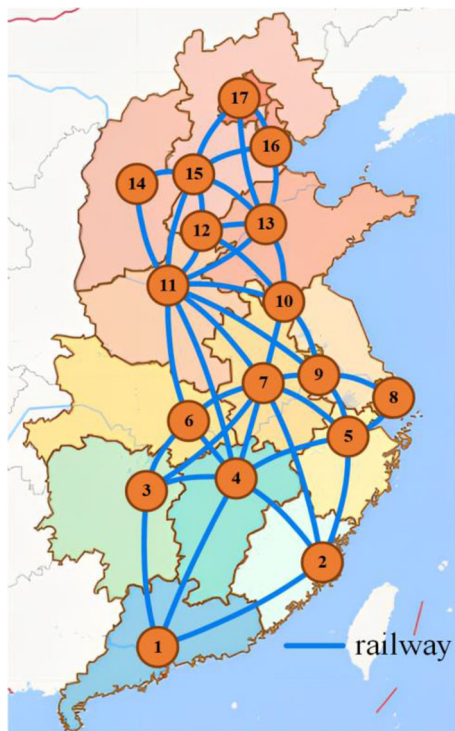
### Directed transportation network

As commonly used algorithms such as PSO, GA, AFO, NSGA-II, and MOPSO algorithms use directed transportation networks, to compare with these algorithms, it is necessary to create corresponding directed transportation networks based on the two transportation networks established in “Transportation network” section.

It is worth noting that changing an undirected network to a directed network requires manual setting of the direction of the routes, which can result in some of the original routes being infeasible. And the lack of certain routes may lead to suboptimal solutions. Therefore, this paper needs to first

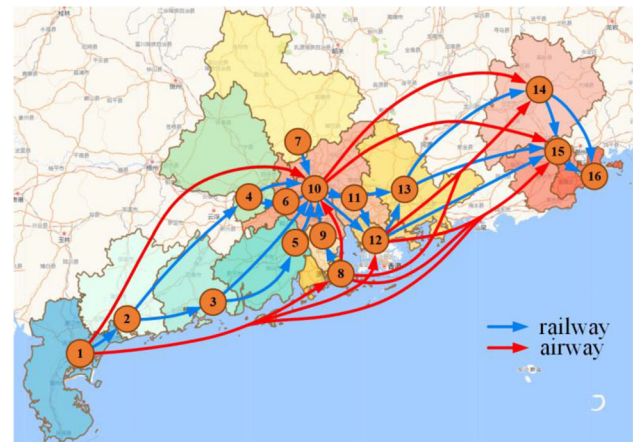
**Table 3** c parameters

Transit	Roadway			Railway			Airway		
	Cost (¥/t)	Time (h/t)	Carbon emission (kgCO <sub>2</sub> /t)	Cost (¥/t)	Time (h/t)	Carbon emission (kgCO <sub>2</sub> /t)	Cost (¥/t)	Time (h/t)	Carbon emission (kgCO <sub>2</sub> /t)
Roadway	–	–	–	8.57	0.268	1.56	11.42	0.2	3.12
Railway	8.57	0.267	1.56	–	–	–	17.14	0.35	6
Airway	11.42	0.2	3.12	17.14	0.35	6	–	–	–

**Fig. 6** Schematic diagram of railway transport

solve for the optimal route using the established undirected network in combination with the proposed method and ensure that the route is feasible in the established directed network. This illustrates, to some extent, the superiority of the established undirected network in finding the optimum, and the superiority of the undirected network in reflecting the cargo transportation behavior, as it does not require manual setting of route directions, providing more route choices and possibilities, allowing the proposed solution to consider the cargo transportation solution more comprehensively and avoiding suboptimal solutions due to manual setting errors.

For the first type of transportation network “[Transportation network](#)” section, the origin of the corresponding directed transportation network is set to 1, and the destination is set to 16. Among them, the highway transport route is shown in Table 10, the transport direction is from the city with the smaller node serial number to the city with the larger

**Fig. 7** Schematic diagram of the directed transportation network of the first network

node serial number, and the railway and airway transport are shown in Fig. 7. In addition, the feasibility of the optimal route is ensured.

For the second transportation network in “[Transportation network](#)” section, the origin of the transport of cargo is set to 1 and the destination is set to 17, and the distances of various transport modes are shown in Table 12. The transport direction is directed from the city with the smaller node serial number to the city with the larger node serial number, and the feasibility of the optimal route is ensured. Among them, railway transportation is shown in Fig. 8.

### Undirected transportation network validation

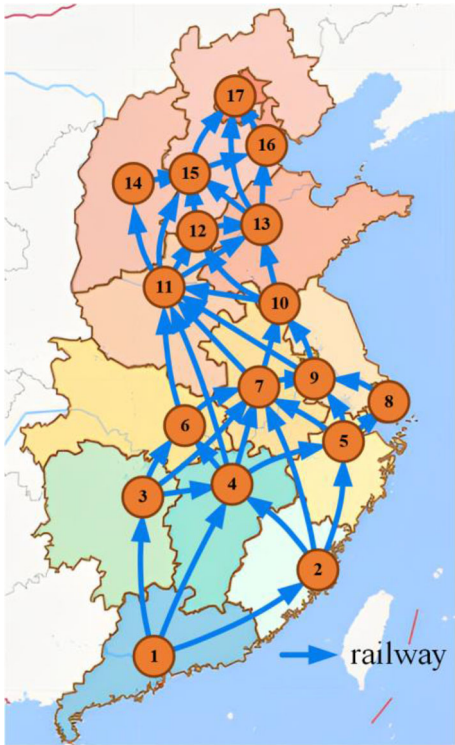
Compared to directed transportation networks, the undirected transportation network built in this paper can reflect the transport behavior of the cargo more accurately and provide more options and possibilities for the algorithm to find the optimal solution. To verify the effectiveness of the undirected transportation networks, directed transportation networks are built based on these networks, with the weight of the cargo set at 20,000 kg, and the GA, PSO, and AFO algorithm is used to find the optimal route with the lowest transport cost.

For the PSO algorithm, according to existing studies, the learning factors were all set to 2.05 [38], the initial inertia

**Table 4** Results comparison of algorithms

	<i>Q</i> -learning	Target value/run time(s)		AFO
		GA	PSO	
1	3082.2/6.075	3811.537/182.994	4025.896/168.955	3082.2/94.166
2	3082.2/6.218	4055.596/207.738	4973.8/172.783	3082.2/96.428
3	3082.2/6.058	4056.591/172.878	3621.794/180.477	3111.9/115.470
4	3082.2/6.032	3766.994/184.51	4025.896/179.884	3894.0/100.253
5	3082.2/6.254	4284.713/164.118	4817.322/187.865	3111.9/101.990
6	3082.2/6.045	<b>3227.4</b> /195.689	4817.322/187.340	3082.2/101.761
7	3082.2/6.195	4313.845/155.428	4817.322/181.526	<b>3082.2</b> /96.781
8	3082.2/6.188	3621.794/186.418	<b>3111.9</b> /199.626	3621.8/106.095
9	3082.2/6.140	3651.494/184.400	4433.594/181.206	3923.7/114.244
10	<b>3082.2</b> /6.140	4284.713/162.631	4269.539/180.042	3894.0/109.211
Average target value	3082.2	3907.4677	4291.439	3388.61
Average run time	6.74	179.681	181.970	106.64

Bold indicates that the value is the minimum value of the target value of the corresponding algorithm



**Fig. 8** Schematic diagram of the directed transportation network of the second network

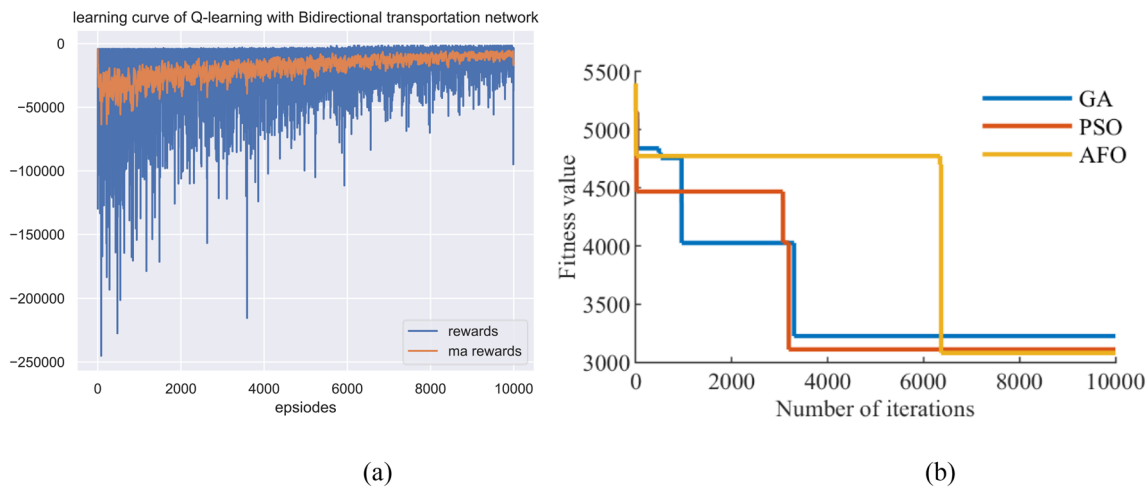
weight in its linear decreasing weight strategy was set to 0.9 and the inertia weight at the maximum count was set to 0.4 [39]. For the GA algorithm, according to existing studies, the crossover probability was set to 0.7 and the variance probability was set to 0.1 [40]. For the AFO algorithm, according to the existing study [41], the minimum trigger interval for the update strategy was set to 5, and the decrease of interval

for the update strategy was set to 2. The number of populations was set to 20 for three algorithms and the maximum number of iterations was set to 10,000. For *Q*-learning,  $\varepsilon_{\text{end}}$  was set to 0.01,  $\varepsilon_{\text{start}}$  was set to 1.0,  $\varepsilon_{\text{decay}}$  was set to 0.9999. The learning rate was set to 0.01 the discount factor  $\gamma$  was set to 0.9 and the maximum number of iterations was also set to 10,000. Ten optimization runs were performed and the optimization results and running times are recorded in Table 4.

In Table 3, the minimum transportation costs obtained by *Q*-learning, GA, PSO, and AFO are 3082.2, 3227.4, 3111.9, and 3082.2, respectively. The iteration curve is shown in Fig. 9. The specific information on the optimal routes is presented in Table 5. Figure 9a shows the *Q*-learning iteration curve; in (a), the rewards curve is the reward obtained by each episode, and the ma rewards curve is the moving average reward. It can be seen that the *Q*-learning algorithm converged to the optimal solution when the iteration reached 8000 times. Due to the  $\varepsilon$ -greedy strategy, the choice of actions may still be random, so there is still a small oscillation in the curve. As can be seen from Fig. 9b, both GA and PSO converged to the optimal solution in around 3000 iterations, but fell into the local optimum at this time, and the global optimal solution was not found in the following more than 6000 iterations. AFO converged to the global optimal solution in around 6000 iterations. In Table 5, only the railway transport mode exists in the optimal route of the three algorithms, but the route obtained by the proposed algorithm passes through fewer nodes and has lower transportation costs.

The average running time of the *Q*-learning combined with the undirected transportation network is only 6.74 s when 10,000 iterations are completed, which is 26 times faster than GA and PSO, and 16 times faster than AFO. The proposed algorithm always solves for the optimal solution, whereas the



**Fig. 9** Convergence curves**Table 5** Comparison of optimal solutions

Algorithm	Optimal path	Cost
<i>Q</i> -learning	1 → 2 → 3 → 10 → 12 → 15 → 16	3082.2
GA	1 → 2 → 3 → 10 → 12 → 13 → 15 → 16	3227.4
PSO	1 → 2 → 3 → 10 → 11 → 13 → 15 → 16	3111.9
AFO	1 → 2 → 3 → 10 → 12 → 15 → 16	3082.2
<i>Q</i> -learning (reverse)	16 → 15 → 12 → 10 → 3 → 2 → 1	3082.2

→ : Highway, → : Railway, → : Airway

GA and PSO algorithms tend to fall into local optima and the average transportation cost of the optimal route found is lower than that of GA, PSO, and AFO. In addition, the undirected transportation network can easily change the origin and destination compared to the directed transportation network. To verify this, we interchanged the origin and the destination, ran the program, and recorded the results. As shown in the last row of Table 5, the resulting optimal route is in the exact opposite direction of the first row and has the same transport costs.

In summary, the multimodal undirected network developed in this paper is effective, more realistic in reflecting the transport of cargo, does not lead to longer search times, combines with *Q*-learning to find optimization results quickly, and facilitates changes to the origin and destination of transport without the need to recreate or modify the network.

**Table 6** Parameter settings of multi-objective *Q*-learning

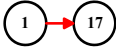

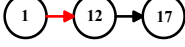
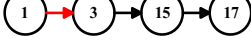

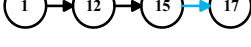




Parameter settings	Value
$\varepsilon$ -strategy	$\varepsilon_{\text{start}}$ 1.0
$(\varepsilon = \varepsilon \cdot \varepsilon_{\text{decay}})$	$\varepsilon_{\text{end}}$ 0.01
	$\varepsilon_{\text{decay}}$ 0.9999
Learning rate $\alpha$	0.01
Discount factor $\gamma$	0.9
Episode	10,000

### Comparison of multi-objective algorithms

For multi-objective *Q*-learning, the model parameter settings are shown in Table 6. Because the cities in the province are close to each other and the transport time of cargo is relatively short, the experimental results may be limited for



**Table 7** Results of multi-objective  $Q$ -learning runs

NO	Weights [cost, time]	Optimal path	Cost (¥)	Carbon cost (¥)	Total cost (¥)	Time (h)
1	[0, 1]		5724.00	91.68	5815.68	3.18
2	[0.01001, 0.98999]		5545.80	82.08	5627.89	6.19
3	[0.01004, 0.98996]		5390.14	76.12	5466.26	7.23
4	[0.01007, 0.98993]		4357.80	35.70	4393.50	14.39
5	[0.0101, 0.9899]		3717.14	11.26	3728.40	17.70
6	[0.01014, 0.98986]		3510.50	10.49	3520.99	21.43
7	[0.01016, 0.98984]		3188.50	8.94	3197.44	25.10
8	[0.06, 0.94]		2346.79	4.73	2351.52	35.90
9	[0.1, 0.9]		2212.28	4.09	2216.37	37.31
10	[1, 0]		1892.55	2.47	1895.02	38.23

→: Highway, →: Railway, →: Airway

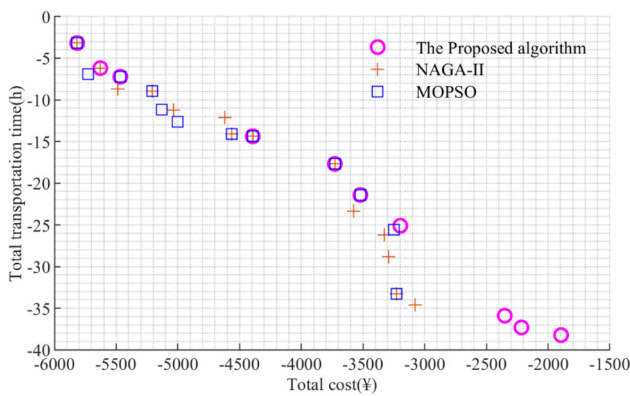
multi-objective reinforcement learning. Moreover, in realistic multimodal transportation, the transportation route of cargo usually spans multiple provinces, and intra-provincial transportation may not be able to fully evaluate the performance of various multimodal transportation strategies. Therefore, to better simulate the real multimodal transportation scenario and better verify the performance of the algorithm. In the experiments in this section, the second transportation network is selected. The origin of the transportation of the cargo is set as 1, the destination is set as 17, and the weight of the cargo is set as 5 000 kg. The Weights [cost, time] are gradually changed from [0, 1] to [1, 0] with a step size of 0.001. Considering that the value of transport cost is two orders of magnitude larger than that of transport time, between [0, 1] and [0.1, 0.9], a step size of 0.00001 is set to explore the impact of weight changes on the algorithm in more detail. In addition, to reduce the operation time, parallel computing is used. The final results obtained are shown in Table 7.

For the customer's delivery options, there are three options: the same day, tomorrow, or the day after tomorrow. If the customer chooses the same-day delivery, then No. 1, No. 2, and No. 3 three options can be chosen. If the customer chooses next-day delivery, then No. 4, No. 5, No. 6, and No. 7

four options can be chosen. If the customer chooses the next-day delivery, then No. 8, No. 9, and No. 10, three options can be chosen.

When the weight of transport cost is 0, only airway transport is selected for the route. When the weight of transport cost is small, the time weight is relatively large, and the possibility of selecting airway transport mode and highway transport mode is higher. As the weight of transport cost gradually increases, the choice of railway transport mode starts to increase. When the weight of cost is 1, only the railway transport mode is selected in the route.

The NSGA-II algorithm and the MOPSO algorithm were selected for comparison. For the NSGA-II algorithm, its crossover rate was set to 0.7, mutation rate to 0.4, mutation probability to 0.02, the initial number of populations to 50, and the number of iterations to 200 [41]. For the MOPSO algorithm, the inertia weight was set to 0.5, the inertia weight damping rate was set to 0.99, the personal learning coefficient was set to 1, and the global learning coefficient was set to 2. For the two algorithms, the initial number of populations is set to 100, and the number of iterations is set to 200. The unidirectional transportation network based on the



**Fig. 10** Comparison of the Pareto front of the algorithm

second transportation network is used. The Pareto front solutions found by the proposed method, NSGA-II algorithm, and MOPSO algorithm are shown in Fig. 10.

The NSGA-II algorithm and the MOPSO algorithm were selected for comparison. For the NSGA-II algorithm, its crossover rate was set to 0.7, mutation rate to 0.4, mutation probability to 0.02, the initial number of populations to 50, and the number of iterations to 200 [42]. For the MOPSO algorithm, the inertia weight was set to 0.5, the inertia weight damping rate was set to 0.99, the personal learning coefficient was set to 1, and the global learning coefficient was set to 2. For the two algorithms, the initial number of populations is set to 100, and the number of iterations is set to 200 [43]. The unidirectional transportation network based on the second transportation network is used. The Pareto front solutions found by the proposed method, NSGA-II algorithm, and MOPSO algorithm are shown in Fig. 10.

From Fig. 10, it can be seen that the number of solutions of the proposed algorithm that are dominated by NSGA-II solutions or MOPSO solutions is 0, while the number of solutions that dominate NSGA-II solutions is 5, and dominate MOPSO solutions is 3. The proposed algorithm in this paper can find more optimal routes when the weight of transport time is larger or the weight of transport cost is larger. The NSGA-II algorithm and the MOPSO algorithm have difficulty in finding the optimal routes when the weight of transport cost is large. For the weights in between, NSGA-II and MOPSO found more solutions than the proposed algorithm, but some of them are dominated by the proposed algorithm's solutions.

To evaluate the set of Pareto solutions obtained by the algorithms, we performed calculations using the HV metrics with the reference point  $(-6000, -40)$ , which is dominated by the solutions of all the algorithms. The computational results show that the HV values of the proposed method, NSGA-II and MOPSO are 73,812, 71,026, and 63,708, respectively. The higher HV values of the proposed method as compared to NSGA-II and MOPSO imply that the proposed method

has a better performance in terms of comprehensive performance.

Regarding the running time of the program, in the study of this paper, for the Pareto frontiers, each solution corresponds to a set of weights (since the weights are close to possibly finding the same solution). Considering the existence of multiple solutions, i.e., multiple weights, for the Pareto frontiers, the model corresponding to each weight needs to be trained. To facilitate the collection of results, the solution is output directly after the training is completed. In other words, for the method proposed in this paper, multiple runs are required to obtain the Pareto frontier, which may result in a longer running time. To cope with this problem, we adopt a strategy of concurrent program execution, where the task is divided into eight programs executed in parallel to fully utilize the number of CPU processing cores. The experimental results show that the running time of the proposed algorithm, NSGA-II, and MOPSO are 2.6 h, 0.18 h, and 0.25 h. However, considering the complexity of the task as well as the value of the results of the algorithm, we believe that such a running time is acceptable [44].

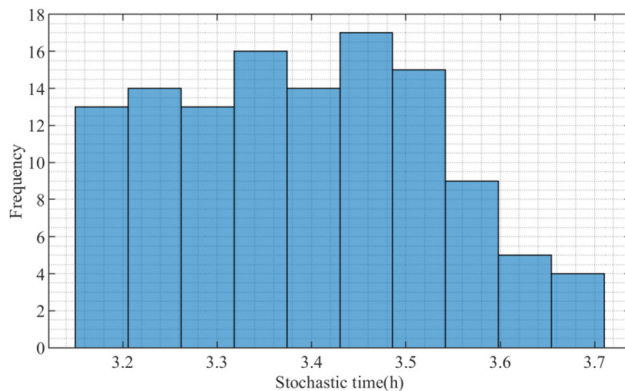
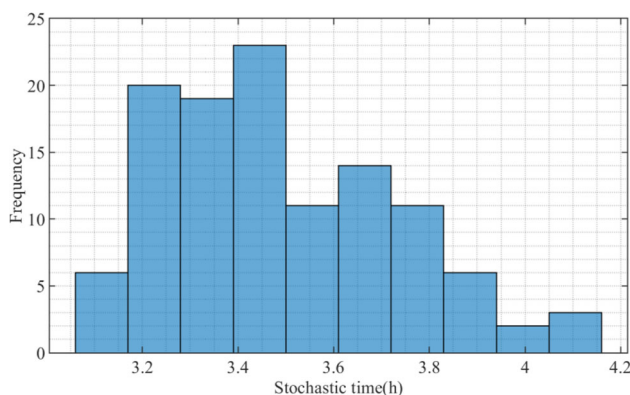
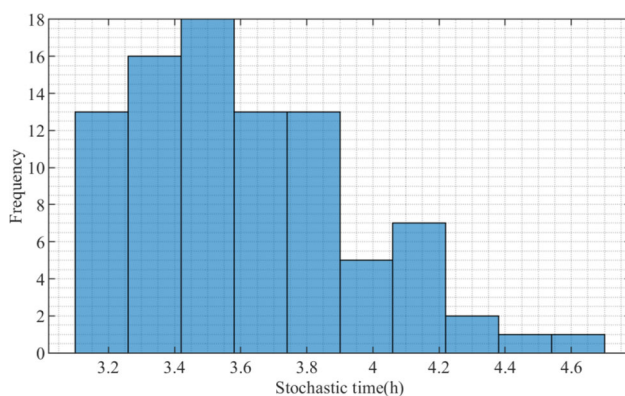
Therefore, after the above analysis, the algorithm proposed in this paper has a better performance in finding the optimal route under multiple weights, especially in the case of a larger weight of transport cost, compared with the NSGA-II algorithm and the MOPSO algorithm, the algorithm proposed in this paper performs better in the multi-objective route optimization problem, but the procedure needs to take more time.

## Time uncertainty

For time uncertainty, a positively skewed normal distribution needs to be established. The ideal transport time (i.e. the time without considering the uncertainty) for each route is set as the mode of the distribution, the skewness is set to 5, and three standard deviations [30] (low, medium, and high standard deviations) are set, taking values of 0.5, 1 and 2. According to the three-sigma rule of normal distribution, a value from probability 0.05–0.73 is selected randomly as the actual transport time of the route. When the weight of transport time is 1, the choice of route is most affected by time, so it is chosen as the study of time uncertainty in this section, and the optimal route is 0–17 and the mode of transport is airway transport. The experimental program for the three standard deviations was run 200 times and the results were recorded as shown in Table 8, where the first row is the deterministic route transport time. The route reliability is calculated as the number of times that the optimal route is still the original deterministic optimal route divided by the total number of the program runs (i.e., 200 times). Figures 11, 12 and 13, respectively, show the histograms of transportation time when the optimal route corresponding to the three standard deviations is still the original deterministic optimal

**Table 8** Comparison of results with different standard deviations

Standard deviation	Average transport time	Route reliability	Percentage increase
–	3.1800	1	0
0.5	3.3880	0.6	6.54%
1	3.4877	0.575	9.68%
2	3.6124	0.445	13.60%

**Fig. 11** Histogram of transport times at a standard deviation of 0.5**Fig. 12** Histogram of transport time at a standard deviation of 1**Fig. 13** Histogram of transport times with a standard deviation of 2

route, and they generally conform to the positively skewed normal distribution.

From Table 8, the average transport time increases for all three standard deviations compared to the deterministic one, which means that the waiting time for the customer increases at this point, which may lead to a decrease in customer service satisfaction. As the standard deviation increases, the reliability of the route decreases because, in some cases, the uncertainty route is no longer optimal and there is a route that takes less time than the original route. Therefore, it is necessary to consider the uncertainty of the transport time. When sufficient historical order data exists, it is necessary to use it to fit probability distributions of transport time and use it to characterize the uncertainty of transport time, which will provide additional benefits to decision-makers.

### Generalization performance

*Q*-learning is a classical reinforcement learning algorithm mainly used for solving problems in finite state and action spaces, where the goal is to learn the optimal policy in a known state and action space. *Q*-learning does not have an explicit notion of generalization performance, as it does not involve learning from finite training data and generalizing to unseen situations. However, in real-world problems, we are often faced with uncertain environments, which necessitates the introduction of the concept of generalization. In this context, generalization ability can be defined as the ability of a trained model to perform well in an uncertain environment. In our previous manuscript, we investigated time uncertainty in “Time uncertainty” section by establishing a positively skewed normal distribution for transport times, setting three standard deviations to investigate the reliability of the original optimal path under different standard deviations. Route reliability is calculated as the number of times the optimal path remains the original deterministic optimal route in the results of the model run under time uncertainty, divided by the total number of the program runs. If the results of the model under uncertainty are the same as the original model, that is to say, the original model is still able to solve the optimal route in this uncertain environment i.e., has good generalization performance. Therefore, the reliability of the original route is equivalent to the generalization performance of the original model in an uncertain environment under this standard deviation. Therefore, according to Table 8, we can get the following Table 9:

The data for reinforcement learning is generated by having the model interact with the environment, and the model is more likely to overfit to the current training environment and thus perform poorly in uncertain environments. As a result, the degree of time uncertainty increases as the standard deviation increases, leading to a decrease in the generalization performance of the proposed method.

**Table 9** Generalization performance of the proposed method

Standard deviation	Generalization performance (%)
–	100
0.5	60
1	57.5
2	44.5

## Conclusion

In multimodal transport route optimization, commonly used algorithms require careful tuning of parameters in their application and suffer from performance degradation when solving large-scale problems. Directed transportation networks result in suboptimal solutions due to manually setting the route in the wrong direction and the need to modify or recreate the network when the origin or destination of cargo transport changes. In actual logistics transportation, cargo transport time is uncertain. To solve the above problems, this paper proposes an optimization framework based on the multi-objective weighted sum  $Q$ -learning algorithm to establish an undirected multiple-node network to better simulate realistic cargo transportation, while facilitating the modification of cargo origin and destination as well as the incorporation of transfer information, and avoid the suboptimal solutions due to incorrectly set route directions manually. Three positively skewed distributions with different standard deviations are established using deterministic transport times as the mode to characterize the time uncertainty. The three objectives of transport cost, transport carbon emission cost, and transport time are studied. The experimental results show that (1) in terms of running time, the proposed algorithm in combination with the established network is 26 times faster than the GA, PSO, and AFO algorithms that use directed transportation networks. In terms of the optimization search results, the proposed algorithm can converge to the optimal solution more stably, while the GA, PSO, and AFO algorithms tend to fall into local optima. In addition, an optimal route can still be searched after modifying the origin and destination of cargo transportation. (2) The corresponding optimal routes are solved according to different weights. Compared with NSGA-II and MOPSO, the proposed algorithm can better solve the boundary of the Pareto front, and dominates the partial solutions of NSGA-II and MOPSO. (3) As the uncertainty increases, the average transport time of the route increases, and at the same time, the reliability of the route decreases, validating the need to consider time uncertainty. However, there are still limitations to the proposed method: the proposed method finds a limited number of solutions in the middle of the Pareto front, which may limit the choice of carriers. Secondly, the weighted sum method

has limitations in some cases [44], to solve the limitations, in the subsequent research, we will carry out an in-depth exploration of the use of a more effective scalar method [45] and compare it with a more excellent multi-objective algorithm [46]. In addition, due to the long computation time of the proposed method, we will study more excellent algorithms [47] to improve our algorithm in the subsequent research.

**Acknowledgements** This work is supported by the Key Research and Development Project of Guangdong Province (Project No. 2021B0101420003).

**Data availability** The authors confirm that the data supporting the findings of this study are available from the following website: for highway transportation distance: <https://www.chinawutong.com/tools/map.html>. For railway transport distances: <http://www.huochepiao.com/>. For airway transport distances: <https://skypearl.csair.com/skypearl/accumulateStandard.html?lang=zh>.

## Declarations

**Conflict of interest** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Appendix

### The first type of transportation network node number and the corresponding city

1-Zhanjiang, 2-Maoming, 3-Yangjiang, 4-Zhaoqing, 5-Jiangmen, 6-Foshan, 7-Qingyuan, 8-Zhuhai, 9-Zhongshan, 10-Guangzhou, 11-Dongguan, 12-Shenzhen, 13-Huizhou, 14-Meizhou, 15-Jieyang, 16-Shantou (Table 10, 11, 12).

### The second transportation network node number and corresponding city

1-Guangzhou, 2-Fuzhou, 3-Changsha, 4-Nanchang, 5-Hangzhou, 6-Wuhan, 7-Hefei, 8-Shanghai, 9-Nanjing, 10-Xuzhou, 11-Zhengzhou, 12-Handan, 13-Jinan, 14-Taiyuan, 15-Shijiazhuang, 16-Tianjin, 17-Beijing.

**Table 10** The highway transportation distance between each node of the first transportation network

Road	Distance	Road	Distance	Road	Distance	Road	Distance	Road	Distance
1-2	96.442	2-12	420.058	4-11	156.486	6-14	413.869	9-14	409.866
1-3	209.328	2-13	452.384	4-12	218.17	6-15	403.611	9-15	385.672
1-4	347.858	2-14	708.75	4-13	231.127	6-16	442.237	9-16	424.188
1-5	356.031	2-15	684.556	4-14	476.192	7-8	204.022	10-11	72.183
1-6	395.283	2-16	723.072	4-15	465.934	7-9	159.034	10-12	138.822
1-7	463.226	3-4	201.393	4-16	504.56	7-10	78.956	10-13	144.913
1-8	407.676	3-5	160.62	5-6	66.736	7-11	130.819	10-14	386.105
1-9	397.637	3-6	199.872	5-7	151.855	7-12	199.617	10-15	375.847
1-10	421.188	3-7	286.473	5-8	82.358	7-13	189.64	10-16	414.473
1-11	456.746	3-8	199.556	5-9	45.409	7-14	397.23	11-12	73.23
1-12	498.9	3-9	202.226	5-10	82.425	7-15	388.384	11-13	90.906
1-13	531.226	3-10	225.777	5-11	115.091	7-16	427.01	11-14	342.179
1-14	787.177	3-11	261.335	5-12	145.598	8-9	43.498	11-15	322.785
1-15	762.983	3-12	303.489	5-13	181	8-10	129.307	11-16	361.301
1-16	801.499	3-13	335.815	5-14	441.392	8-11	130.628	12-13	88.354
2-3	130.486	3-14	591.921	5-15	417.198	8-12	159.958	12-14	341.203
2-4	264.024	3-15	567.727	5-16	455.714	8-13	196.148	12-15	317.265
2-5	277.189	3-16	606.243	6-7	96.083	8-14	452.221	12-16	334.459
2-6	316.441	4-5	101.193	6-8	127.767	8-15	428.027	13-14	259.784
2-7	379.392	4-6	89.046	6-9	81.927	8-16	466.543	13-15	249.526
2-8	328.834	4-7	123.151	6-10	34.85	9-10	83.467	13-16	288.152
2-9	318.795	4-8	183.56	6-11	95.209	9-11	88.921	14-15	109.188
2-10	342.346	4-9	145.768	6-12	139.432	9-12	120.186	14-16	155.811
2-11	377.904	4-10	99.172	6-13	172.677	9-13	153.793	15-16	53.389

**Table 11** The railway and airway transport distances between the nodes of the first transportation network

Road	Distance	Road	Distance	Road	Distance
1-2	(93, −)	5-10	(65, −)	10-15	(−, 353)
1-8	(−, 353)	6-10	(22, −)	11-12	(61, −)
1-10	(−, 369)	7-10	(83, −)	11-13	(54, −)
1-12	(−, 398)	8-9	(46, −)	12-13	(56, −)
1-15	(−, 697)	8-10	(−, 115)	12-14	(−, 430)
2-3	(117, −)	8-14	(−, 696)	12-15	(337, 363)
2-4	(477, −)	8-15	(−, 428)	13-14	(286, −)
3-5	(158, −)	9-10	(70, −)	13-15	(325, −)
3-10	(223, −)	10-11	(71, −)	14-15	(110, −)
4-6	(87, −)	10-12	(104, −)	14-16	(171, −)
4-10	(109, −)	10-14	(−, 322)	15-16	(60, −)



**Table 12** The distance of three transportation modes between each node of the second transportation network (highway, railway, airway)

Road	Distance	Road	Distance	Road	Distance	Road	Distance
1-2	(869.693, 861, 691)	3-7	(701.808, 722, 641)	5-16	(1123.095, –, 1133)	9-12	(800.029, –, 648)
1-3	(669.378, 707, 562)	3-8	(1045.236, –, 895)	5-17	(1250.238, –, 1135)	9-13	(617.181, –, 579)
1-4	(777.857, 948, 667)	3-9	(853.752, –, 702)	6-7	(385.173, 339, 325)	9-14	(1087.451, –, 974)
1-5	(1249.828, –, 1056)	3-10	(996.033, –, 784)	6-8	(835.498, –, 676)	9-15	(897.081, –, 857)
1-6	(978.105, –, 823)	3-11	(803.305, –, 768)	6-9	(539.414, –, 459)	9-16	(888.566, –, 907)
1-7	(1207.947, –, 1047)	3-12	(1038.465, –, –)	6-10	(648.088, –, 501)	9-17	(1015.709, –, 926)
1-8	(1432.588, –, 1198)	3-13	(1177.306, –, 1013)	6-11	(509.625, 536, 480)	10-11	(371.934, 349, –)
1-9	(1356.018, –, 1121)	3-14	(1234.142, –, 1064)	6-12	(731.947, –, –)	10-12	(488.124, 596, –)
1-10	(1527.686, –, 1281)	3-15	(1203.402, –, 1249)	6-13	(854.323, –, 729)	10-13	(319.856, 319, 275)
1-11	(1445.758, –, 1291)	3-16	(1449.043, –, 1353)	6-14	(951.637, –, 866)	10-14	(769.35, –, 607)
1-12	(1680.627, –, 1519)	3-17	(1478.469, –, 1363)	6-15	(894.327, –, 872)	10-15	(585.868, –, –)
1-13	(1819.759, –, 1544)	4-5	(522.549, 582, 461)	6-16	(1126.06, –, 998)	10-16	(608.256, –, –)
1-14	(1876.595, –, 1624)	4-6	(342.064, 344, 343)	6-17	(1169.394, –, 1087)	10-17	(709.443, –, 676)
1-15	(1845.855, –, 1660)	4-7	(431.738, 462, 450)	7-8	(466.113, –, 390)	11-12	(252.006, 247, –)
1-16	(2091.496, –, 1816)	4-8	(695.019, –, 596)	7-9	(169.844, 173, 145)	11-13	(446.44, 646, 359)
1-17	(2120.922, –, 1908)	4-9	(583.682, –, 462)	7-10	(324.512, 295, 362)	11-14	(446.977, 617, 346)
2-3	(847.955, –, 675)	4-10	(750, –, 620)	7-11	(565.899, 602, 486)	11-15	(417.173, 408, 421)
2-4	(556.383, 547, 451)	4-11	(842.645, 1152, 860)	7-12	(745.97, –, –)	11-16	(694.162, –, 583)
2-5	(613.012, 681, 488)	4-12	(1050.135, –, –)	7-13	(640.007, –, 558)	11-17	(692.24, –, 646)
2-6	(887.838, –, 700)	4-13	(1065.497, –, 895)	7-14	(1013.744, –, 808)	12-13	(259.476, 283, –)
2-7	(868.294, 808, 704)	4-14	(1281.119, –, 1090)	7-15	(874.018, –, 736)	12-14	(313.025, –, –)
2-8	(768.971, –, 609)	4-15	(1210.59, –, 1051)	7-16	(928.85, –, 807)	12-15	(168.417, 161, –)
2-9	(869.093, –, 665)	4-16	(1328.454, –, 1169)	7-17	(1017.166, –, 918)	12-16	(445.37, –, –)
2-10	(1170.587, –, 911)	4-17	(1404.13, –, 1270)	8-9	(298.882, 295, 255)	12-17	(443.453, –, –)
2-11	(1388.419, –, 1113)	5-6	(721.577, –, 568)	8-10	(582.639, –, 494)	13-14	(518.86, –, –)
2-12	(1556.881, –, –)	5-7	(415.012, 627, 327)	8-11	(942.007, –, 815)	13-15	(313.9, 904, –)
2-13	(1481.624, –, 1209)	5-8	(173.321, 159, 176)	8-12	(1019.009, –, 888)	13-16	(312.877, 357, –)
2-14	(1829.611, –, 1451)	5-9	(278.554, 442, 227)	8-13	(813.217, –, 736)	13-17	(401.531, 410, 412)
2-15	(1683.902, –, 1450)	5-10	(599.354, –, 486)	8-14	(1324.416, –, 1082)	14-15	(223.086, 210, –)
2-16	(1755.016, –, 1470)	5-11	(932.294, –, 787)	8-15	(1119.456, –, 1005)	14-16	(518.018, –, 440)
2-17	(1862.092, –, 1586)	5-12	(1059.366, –, 1250)	8-16	(1081.464, –, 953)	14-17	(489.811, –, 431)
3-4	(336.642, 342, 331)	5-13	(852.498, –, 772)	8-17	(1208.607, –, 1088)	15-16	(310.848, 429, 279)
3-5	(872.766, –, 744)	5-14	(1360.82, –, 1084)	9-10	(325.54, 348, –)	15-17	(290.687, 281, 364)
3-6	(335.652, 362, 288)	5-15	(1155.86, –, 1011)	9-11	(659.568, 697, 568)	16-17	(136.833, 127, 180)

## References

- Bontekoning YM, Macharis C, Trip JJ (2004) Is a new applied transportation research field emerging? A review of intermodal rail–truck freight transport literature. *Transp Res Part A Policy Pract* 38(1):1–34
- Macharis C, Bontekoning YM (2004) Opportunities for OR in intermodal freight transport research: a review. *Eur J Oper Res* 153(2):400–416
- Bortolini M, Faccio M, Ferrari E et al (2016) Fresh food sustainable distribution: cost, delivery time and carbon footprint three-objective optimization. *J Food Eng* 174:56–67
- Bauer J, Bektas T, Crainic TG (2010) Minimizing greenhouse gas emissions in intermodal freight transport: an application to rail service design. *J Oper Res Soc* 61(3):530–542
- Zheng CJ, Sun K, Gu YH et al (2022) Multimodal transport path selection of cold chain logistics based on improved particle swarm optimization algorithm. *J Adv Transp* 2022:1
- Resat HG, Turkay M (2015) Design and operation of intermodal transportation network in the Marmara region of Turkey. *Transp Res E Log* 83:16–33
- Zhang H, Li Y, Zhang QP et al (2021) Route selection of multimodal transport based on China railway transportation. *J Adv Transp* 2021:1

8. Jiang J, Zhang D, Meng Q et al (2020) Regional multimodal logistics network design considering demand uncertainty and CO<sub>2</sub> emission reduction target: a system-optimization approach. *J Clean Prod* 2020:248
9. Fazayeli S, Eydi A, Kamalabadi IN (2018) Location-routing problem in multimodal transportation network with time windows and fuzzy demands: presenting a two-part genetic algorithm. *Comput Ind Eng* 119:233–246
10. Liu H, Song G, Liu T et al (2022) Multitask emergency logistics planning under multimodal transportation. *Mathematics* 10(19):1
11. Xu D, Wenfeng L, Lanbo Z (2013) Ant colony optimisation for a resource-constrained shortest path problem with applications in multimodal transport. *Int J Model Ident Control* 18(3):268–275
12. Zhang Q, Wu K, Shi Y (2020) Route planning and power management for PHEVs with reinforcement learning. *IEEE Trans Veh Technol* 69(5):4751–4762
13. Xu Y, Fang M, Chen L et al (2022) Reinforcement learning with multiple relational attention for solving vehicle routing problems. *IEEE Trans Cybern* 52(10):11107–11120
14. Feng S, Duan P, Ke J et al (2022) Coordinating ride-sourcing and public transport services with a reinforcement learning approach. *Transp Res Part C Emerg Technol* 138:1
15. Hu R, Xu J, Chen B et al (2020) TAP-net: transport-and-pack using reinforcement learning. *ACM Trans Graph* 39(6):1
16. Watkins CJCH, Dayan P (1992) Technical note: *Q*-learning. *Mach Learn* 8(3):279–292
17. Jaakkola T, Jordan MI, Singh SP (1993) Convergence of stochastic iterative dynamic programming algorithms. In: *Proceedings of the 6th international conference on neural information processing systems*, pp 703–710
18. Tsitsiklis JN (1994) Asynchronous stochastic approximation and *Q*-learning. *Mach Learn* 16(3):185–202
19. Baird L (1995) Residual algorithms: reinforcement learning with function approximation. *Machine learning*. In: *Proceedings of the 12th international conference on machine learning*, pp 30–37
20. Liu C, Xu X, Hu D (2015) Multiobjective reinforcement learning: a comprehensive overview. *IEEE Trans Syst Man Cybern Syst* 45(3):385–398
21. Hayes CF, Radulescu R, Bargiacchi E et al (2022) A practical guide to multi-objective reinforcement learning and planning. *Autonomous Agents Multiagent Syst* 36(1):1
22. Ngai DCK, Yung NHC (2011) A multiple-goal reinforcement learning method for complex vehicle overtaking maneuvers. *IEEE Trans Intell Transp Syst* 12(2):509–522
23. Zhao Y, Chen Q, Hu W et al (2010) Multi-objective reinforcement learning algorithm for MOSDMP in unknown environment. In: *8th world congress on intelligent control and automation (WCICA)*, pp 3190–3194
24. Vamplew P, Dazeley R, Berry A et al (2011) Empirical evaluation methods for multiobjective reinforcement learning algorithms. *Mach Learn* 84(1–2):51–80
25. Zeng F, Zong Q, Sun Z et al (2010) Self-adaptive multi-objective optimization method design based on agent reinforcement learning for elevator group control systems. In: *8th world congress on intelligent control and automation (WCICA)*, pp 2577–2582
26. Zhang X, Jin F-Y, Yuan X-M et al (2021) Low-carbon multimodal transportation path optimization under dual uncertainty of demand and time. *Sustainability* 13(15):1
27. Sun Y (2020) Fuzzy approaches and simulation-based reliability modeling to solve a road–rail intermodal routing problem with soft delivery time windows when demand and capacity are uncertain. *Int J Fuzzy Syst* 22(7):2119–2148
28. Ramezani M, Bashiri M, Tavakkoli-Moghaddam R (2013) A new multi-objective stochastic model for a forward/reverse logistic network design with responsiveness and quality level. *Appl Math Model* 37(1–2):328–344
29. Demir E, Burgholzer W, Hrusovsky M et al (2016) A green intermodal service network design problem with travel time uncertainty. *Transp Res Part B Methodol* 93:789–807
30. Juan A, Faulin J, Grasman S et al (2011) Using safety stocks and simulation to solve the vehicle routing problem with stochastic demands. *Transp Res Part C Emerg Technol* 19(5):751–765
31. Peng Y, Yong P, Luo Y (2021) The route problem of multimodal transportation with timetable under uncertainty: multi-objective robust optimization model and heuristic approach. *Rairo Oper Res* 55:S3035–S3050
32. Baykasoglu A, Subulan K (2019) A fuzzy-stochastic optimization model for the intermodal fleet management problem of an international transportation company. *Transp Plan Technol* 42(8):777–824
33. Haddadsisakht A, Ryan SM (2018) Closed-loop supply chain network design with multiple transportation modes under stochastic demand and uncertain carbon tax. *Int J Prod Econ* 195:118–131
34. Sun Y, Liang X, Li X et al (2019) A fuzzy programming method for modeling demand uncertainty in the capacitated road-rail multimodal routing problem with time windows. *Symmetry* 11(1):91
35. Farahani A, Genga L, Dijkman R et al (2021) Online multimodal transportation planning using deep reinforcement learning. In: *IEEE international conference on systems, man, and cybernetics (SMC)*, pp 1691–1698
36. Barron EN, Ishii H (1989) The Bellman equation for minimizing the maximum cost. *Nonlinear Anal Theory Methods Appl* 13(9):1067–1090
37. Roijers DM, Vamplew P, Whiteson S et al (2013) A survey of multi-objective sequential decision-making. *J Artif Intell Res* 48:67–113
38. Cao B, Sun K, Li T et al (2018) Trajectory modified in joint space for vibration suppression of manipulator. *IEEE Access* 6:57969–57980
39. Yang Y, Xu H-Z, Li S-H et al (2022) Time-optimal trajectory optimization of serial robotic manipulator with kinematic and dynamic limits based on improved particle swarm optimization. *Int J Adv Manuf Technol* 120(1–2):1253–1264
40. Zhai L, Feng S (2022) A novel evacuation path planning method based on improved genetic algorithm. *J Intell Fuzzy Syst* 42(3):1813–1823
41. Yang Z, Deng L, Wang Y et al (2021) Aptenodytes Forsteri optimization: algorithm and applications. *Knowl Based Syst* 2021:232
42. Zobaa AF (2019) Mixed-integer distributed ant colony multi-objective optimization of single-tuned passive harmonic filter parameters. *IEEE Access* 7:44862–44870
43. Thabit S, Mohades A (2019) Multi-robot path planning based on multi-objective particle swarm optimization. *IEEE Access* 7:2138–2147
44. Wang Z, Zhen H-L, Deng J et al (2022) Multiobjective optimization-aided decision-making system for large-scale manufacturing planning. *IEEE Trans Cybern* 52(8):8326–8339
45. Zheng R, Wang Z (2023) A generalized scalarization method for evolutionary multi-objective optimization. *Proc AAAI Conf Artif Intell* 37:12518–12525
46. Wang Z, Zhang Q, Zhou A et al (2016) Adaptive replacement strategies for MOEA/D. *IEEE Trans Cybern* 46(2):474–486
47. Li K, Zhang T, Wang R (2021) Deep reinforcement learning for multiobjective optimization. *IEEE Trans Cybern* 51(6):3103–3114

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.