IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 8 April 2024, accepted 10 April 2024, date of publication 16 April 2024, date of current version 26 April 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3389749

RESEARCH ARTICLE

Energy-Efficient Dynamic Vehicle Routing for Waste Collection by Q-Learning-Based Hyperheuristic Particle Swarm Optimization

YUN ZHAO[1], XIAONING SHEN[2,3,4], WENYAN CHEN[2], AND HONGLI PAN[2]

[1]School of Electrical Engineering, Nanjing Vocational University of Industry Technology, Nanjing, Jiangsu 210023, China
[2]School of Automation, Nanjing University of Information Science and Technology, Nanjing, Jiangsu 210044, China
[3]Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology (CICAEET), Nanjing University of Information Science and Technology, Nanjing, Jiangsu 210044, China
[4]Jiangsu Key Laboratory of Big Data Analysis Technology, Nanjing University of Information Science and Technology, Nanjing 210044, China

Corresponding author: Xiaoning Shen (sxnystsyt@sina.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61502239, and in part by the Natural Science Foundation of Jiangsu under Grant BK20150924.

ABSTRACT During the process of waste collection, various unpredicted disturbances might occur. Meanwhile, vehicle transport is an important source of carbon emissions. In this study, an energy-efficient multi-trip dynamic vehicle routing model is established for waste collection, which introduces two types of dynamic events: new collecting requirements of waste sites and vehicle breakdowns. A Q-learning-based hyperheuristic particle swarm optimization (QLHPSO) is proposed as a dynamic rescheduling method to solve the model. A set of low-level heuristics (LLHs) are designed by combining the learning operators in particle swarm optimization and local search operators. A Q-learning-based high-level strategy is developed to find a suitable LLH for each evolutionary state based on historical performance of LLHs. When rescheduling is triggered, a response mechanism is incorporated to construct initial population by utilizing the features of dynamic events and historical elites. Extensive experimental results on one real instance and nine synthetic instances show that QLHPSO can react to the environmental changes rapidly, and reschedule the vehicle routes with lower cost and carbon emissions compared to the state-of-the-art algorithms.

INDEX TERMS Waste collection, dynamic scheduling, energy-efficient, hyperheuristic, Q-learning.

I. INTRODUCTION

With the growth of global population and acceleration of urbanization, the population gradually migrate to the city and the living environment has been changed greatly, which lead to more and more household waste. For instance, according to the estimates released by the Waste Engineering Institute of Japan in 2020, the amount of waste produced in the world will reach 32 billion tons a year by 2050, 4.2 times the amount in 2000. The accumulation of waste results in the rapid deterioration of the global environment and rise of the carbon emissions. It was reported by International Energy Agency (IEA) that global carbon emissions had reached 33.143 billion tons during 2018, up 1.7 percent from the previous year. Facing such a huge challenge, the world attaches great importance to it and many countries begin to implement the project of Integrated Solid Waste Management (ISWM) [1]. ISWM includes generation of waste, source-separation, storage, collection, processing and disposal [2], among which waste collection plays a role of bridge linking the fore-end waste generation with the terminal waste treatment [3]. Waste collection typically involves garbage vehicles starting from the depot and traveling in certain routes to collect waste by visiting all waste sites, which cost a large amount of budget. For example, the annual waste management cost in the United States is about

The associate editor coordinating the review of this manuscript and approving it for publication was Diego Oliva.

$ 20 billion, of which waste collection accounts for more than half [4]. Furthermore, road traffic is a main contributor to the large amount of daily carbon emissions. In particular, heavy vehicles like garbage trucks take up a small part of the vehicle population but a huge proportion of tailpipe emissions [5]. In view of the problems caused by waste, the rational energy-efficient vehicle routing is significant to ISWM considering both the environmental protection and economy.

So far, a lot of scholars have paid attentions to the waste collection problem [6]. However, most of the related work only studied the static vehicle routing. It was assumed that the environmental information like the number of waste sites and vehicle conditions can be obtained in advance and remain unchanged. However, during actual waste collection, various dynamic disturbances might occur, such as new collecting requirements of waste sites, sudden vehicle breakdowns and so on, which would lead to alterations of the waste collection environment. When dynamic events happen, the original best found route obtained based on static approaches might be no longer the optimal in the new environment, which might even cause the infeasible routes. To our knowledge, dynamic vehicle routing for waste collection has been scarcely reported in the existing literature. Thus, it would be significant to consider waste collection as a dynamic vehicle routing problem to tackle changing environments, and design a suitable vehicle routing approach to cater for environmental variations and decrease the influence of random events on transportation cost and carbon emission. In addition, most existing studies model waste collection as a single-trip vehicle routing problem [6]. Nevertheless, the use of each vehicle requires driver costs and maintenance costs in real life. To fully utilize the driver and vehicle resources, each vehicle must take multiple trips to collect waste during the driver's working hours. To address these problems, this paper aims for constructing an energy-efficient multi-trip dynamic vehicle routing model for waste collection (EMDVRWC).

Vehicle routing problem (VRP) has been proved to be an NP-hard problem [7]. Exact algorithms are difficult to generate satisfactory solutions within a limited time span for the medium to large scale NP-hard problem [8]. Heuristic methods are able to quickly construct solutions but incapable of guaranteeing the solution quality for the lack of global information [9]. Meta-heuristics are search-based approaches inspired by the social behavior of biological population. They are more suitable for the NP-hard problems since they can produce good solutions in an acceptable computational time. Optimization algorithms have substantial effect on the quality of strategies [10]. Particle swarm optimization (PSO) [11] is one of the meta-heuristics which simulates the process of birds randomly searching for food. PSO has the capabilities of memory, self-learning and social learning. Thus, PSO is known for rapid searching and high precision, and has been widely used in different types of VRPs [12], [13], [14]. Apparently, a variety of velocity and position update methods (i.e., learning operators) have been designed in these PSO research works. Refer to the no free lunch theorem, determining the operator ideally suited for a specific vehicle routing model usually requires numerous numerical simulations. Nevertheless, the customer demands and the vehicle conditions might change with time, leading to variations in the decision space dimension and the fitness landscape. The learning operator that performs well in a particular environment might deteriorate when facing another case, which urges for an improvement in adaptability to dynamic vehicle routing problems for waste collection.

With the aim of automatically identifying the appropriate operators with the progress of evolution, the hyperheuristic algorithm (HHA) was proposed [15]. HHA provides a high-level strategy (HLS) to manipulate a group of Low-Level Heuristics (LLHs). For its self-adaptation to any environmental state without hand regulation, HHA has been studied by a number of scholars [16]. According to the characteristics of the search space, HHA can be classified into heuristic selection and heuristic generation. Among them, heuristic selection chooses an existing LLH suitable for the current state by HLS, which tries to improve the current solution through iteration until the termination condition is met. Heuristic generation generates a new heuristic algorithm from the components of existing LLHs, which is more flexible but its design and use are more complex.

Since the considered problem EMDVRWC has a dynamically changing environment, with the purpose of improving the robustness of HHA to different disturbances, a commonly used value-based reinforcement learning algorithm named Q-learning [17] is adopted in this paper as HLS to select the ideal LLH. Based on the instant reward of a state-action pair, the long-term reward preserved in Q-table is updated [18]. Through the iterative training of Q-learning, the environment with different states gradually learn to choose the most suitable search operators for themselves. There are three main technique issues of Q-learning, including divisions of the environmental states, design of right actions and an update strategy for reward. Most existing literature on Q-learning-based hyperheuristics divided the states based on the time varying landscape [19], [20], [21]. According to such a state partition, a heuristic that is beneficial to increase diversity and enhance exploration might not be selected as the search operator, and premature convergence tends to occur. Therefore, a new state division method considering both the convergence and diversity needs to be devised. Moreover, based on both the problem domain and the tasks of different evolutionary stages, elaborate design of a set of LLHs is required.

To solve the above problems, this study investigates a predictive-rescheduling method based on Q-learning-based hyperheuristic particle swarm optimization (QLHPSO) for the energy-efficient multi-trip vehicle routing problem of waste collection, with the aim of identifying the most

appropriate LLH with the progress of evolution considering the dynamic features of waste collection.

The main contributions are summarized as follows: (1) in order to address the dynamic features of VRP in waste collection, two kinds of dynamic events including new collecting requirements of waste sites and vehicle breakdowns are introduced. Both the carbon emissions and the transportation cost are minimized subjecting to the practical constraints. Besides, to improve the resource utilization, each vehicle is allowed to have multiple trips to collect waste, (2) with the aim of reacting to the environmental changes timely and promote convergence, a heuristic population initialization strategy is presented by utilizing the characteristics of dynamic events and memorizing the historical optimum, and (3) eight LLHs are designed to enrich the search mode of particles by combining four learning operators and two enhanced local search operators, which focus on different search functions. By defining states considering both convergence and diversity, Q-learning-based HLS is developed to determine an appropriate LLH under each state.

The remainder of the paper is organized as follows. Section **II** reviews the related work. Section **III** formulates the mathematical model of EMDVRWC. The proposed algorithm for solving the model is described in Section **IV**. Section **V** discusses experimental studies. The paper is concluded in Section **VI**.

II. RELATED WORK

The notion of waste collection was first introduced by Beltrami and Bodin [22]. Since then, a lot of promising results have been achieved, which can mainly be divided into two aspects: problem modeling and the approaches to solve the problem.

A. MODELS OF THE WASTE COLLECTION PROBLEM

In order to make the model more realistic, some studies incorporate various practical factors during the waste collection into the model. Others attach much importance to the environmental and ecological protection.

With regard to the practicability, Louati et al. [23] studied a generalized waste collection vehicle routing problem, which contained multiple transfer stations and time windows. Mat et al. [24] constructed a vehicle routing model for waste collection with a variable travel speed. Tirkolaee et al. [25] considered uncertainties in waste collection, including uncertain amount of waste and variations of the vehicle capacity. Shi et al. [26]studied the waste collection problem with multiple waste depots. Zafor et al. [27] applied sensor nodes to monitor the collection of dry and wet waste separately and determine if the dustbin is full or not. Shen et al. [28] studied the model based on a route-related uncertain set to tackle the uncertainty of customer demand and the weight-related energy consumption of electronic vehicles. Tee and Cruz [29] developed a model that extended the current developments of VRP by considering wait times in specific nodes.

With the increasing emphasis of human society on the ecological environment, more and more scholars consider the energy-efficient related objectives besides the economy objectives like transportation cost, time or distance. Tirkolaee et al. [25] modeled the green capacitated arc routing problem (CARP), which aimed to minimize the sum of the greenhouse gas emission cost and the vehicle operation cost. Molina et al. [30] took eco-efficiency as a performance indicator to design a waste collection routing model for a single landfill to reduce carbon emissions caused by waste collection. Wu et al. [31] formulated a model for the green vehicle routing problem, taking the cost of greenhouse gas emission into account. In Erdem [32], a fleet of heterogeneous electric vehicles were assigned to carry out a number of visits to the places to reduce the total travel costs and harmful gas emissions. Pamukçu et al. [33] developed a model based on Tier-I method in which greenhouse gas emissions from waste collection and transport have been calculated. Hu et al. [34] incorporated the lifecycle carbon emissions of fuel and discussed the economic and environmental impacts of vehicle idling.

Despite the existing studies on waste collection have introduced some uncertainties into the waste collection model, occurrences of dynamic events have not been considered. Moreover, most existing models on waste collection are based on the VRP model with a single trip. To cover the above shortages, we formulate an energy-efficient multi-trip dynamic vehicle routing model, which is different from the previous work in that: (1) two kinds of practical dynamic events, including new collecting requirements of waste sites and vehicle breakdowns, occur during the waste collection one by one following a Poisson distribution; and (2) more practical factors such as multiple trips of each vehicle and the maximum working hours of the drivers are incorporated.

B. APPROACHES TO SOLVE THE WASTE COLLECTION PROBLEM

The VRP in waste collection involves two strongly coupled subproblems: 1) vehicle selection referring to select the minimal number of vehicles to work together on the waste collection tasks; and 2) finding the most appropriate route for each selected vehicle. It has been proved to be an NP-hard problem [35], for which exact algorithms cannot get the optimal solution within polynomial time for medium to large-scale problems. Some heuristic algorithms have been applied to this problem. Cortinhal et al. [36] adopted a two-stage heuristic algorithm to optimize the household waste collection. Akhtar et al. [37] presented a modified backtracking search algorithm for the capacitated vehicle routing problem (CVRP) with the concept of smart bin. Louati et al. [23] proposed a heuristic smart routing algorithm for municipal solid waste collection. Shi et al. [26] solved the multi-depot VRP in waste collection using a sector-scanning algorithm. Jin et al. [38] proposed a single-solution based intelligent heuristic search algorithm to solve an arc-routing problem

with time-dependent penalty cost. Marseglia et al. [39] solved the vehicle routes in waste collection in an optimal way by developing an adaptive algorithm of overflow deviated to the immediate neighborhood.

Meta-heuristic algorithms search on one or a set of solutions iteratively to get the new approximate solutions, which allows exploring a larger scope of the decision space [40] compared to heuristic algorithms. In recent years, more and more scholars have applied meta-heuristics to the waste collection. Tirkolaee et al. [41] adopted a hybrid simulated annealing algorithm based on an efficient cooling equation to handle the CARP for waste collection. Hannan et al. [42] developed a modified PSO in CVRP which could determine the optimal vehicle routing plan for waste collection. Wei et al. [43] presented an artificial bee colony algorithm to handle a waste collection problem with the midway disposal pattern. Wichapa et al. [44] designed a hybrid genetic algorithm to solve the infectious waste collection vehicle routing model. Qiao et al. [45] used a two-stage algorithm, which included PSO and tabu search to solve the CVRP. Tomitagawa et al. [46] presented an adapted ant colony optimization algorithm to find the energy-efficient paths of the waste collection robots.

Despite various meta-heuristic algorithms have been applied to the waste collection problem, there are still some open issues, e.g., the fixed search operators are unable to adapt well to the environmental or model changes, and might lead to a lack of diversity when creating the offspring individuals. To solve the formulated dynamic model and overcome the problems with the existing algorithms, a predictive-rescheduling method based on a Q-learning-based hyperheuristic particle swarm optimization is designed. This approach differs from the previous ones in that: (1) a dynamic response mechanism is developed by utilizing the characteristics of dynamic events and memorizing the historical optimum, (2) eight low-level heuristics (LLHs) are designed by combining four learning operators with different search tasks and two enhanced local search operators, and (3) a Q-learning based high-level strategy is adopted to control the LLHs, where the population with different states gradually learn to choose the ideal operators through the training of Q-learning.

III. MATHEMATICAL MODELING OF EMDVRWC

This section gives a description of the considered problem EMDVRWC and constructs a mathematical model for it.

A. DESCRIPTIONS OF EMDVRWC

EMDVRWC can be defined as a directed network G = (V, E), where V is the set of points including the depot, disposal station and waste sites, and E = $\{(i, j)| i, j \in V\}$ is an edge set. With the aim of minimizing both the carbon emissions and the transportation cost, EMDVRWC involves selecting the minimal number of vehicles and finding the most appropriate route for each vehicle, subjecting to the constraints of drivers' working hours and the vehicle capacity. Assume that all the vehicles are homogeneous with both the maximum capacity of each vehicle and the maximum working hours of each driver being equal. Each vehicle is allowed to have multiple trips, the number of which is decided by the maximum available time and the capacity of vehicles. In order to address the dynamic characteristics of EMDVRWC, two types of dynamic events that often occur in waste collection are introduced, which are described as follows.

1) DYNAMIC EVENT 1: NEW COLLECTING REQUIREMENTS OF WASTE SITES

During the waste transportation, the waste collection company might temporarily receive the new cleaning task from the waste sites other than the ones they are previously responsible for or a large amount of waste is suddenly added to the sites that have been served. In this case, the transportation cost and carbon emissions are likely to be increased if the original route is still used. It needs to adjust the vehicle routes through dynamic rescheduling after the occurrence of such dynamic events.

2) DYNAMIC EVENT 2: VEHICLE BREAKDOWNS

Due to the lack of vehicle maintenance or complicated road conditions, a vehicle might break down in transit. As a result, the vehicle cannot continue to serve the following communities. At this time, collecting the waste along the old route would become infeasible, and a new vehicle route should be obtained by dynamic rescheduling.

B. MATHEMATICAL MODELLING OF EMDVRWC

This section establishes the mathematical model of the considered problem EMDVRWC.

1) NOTATIONS

The notations in the EMDVRWC model are listed in Table 1.

2) MATHEMATICAL MODELLING

Assume the initial time is t_0. For any scheduling point $t_l \geqslant t_0$, considering all the current information gathered from the environment, which includes the waste sites to be collected, a set of available vehicles with departure time, the remaining capacity and the last visited site before t_l, EMDVRWC consists in selecting the minimum number of vehicles and determining the most appropriate route for each selected vehicle (multiple trips are allowed in each route) by optimizing the objective subjecting to a group of constraints:

$$\begin{aligned}
&\min C_{\text{total}}(t_l) \\
&= C_{\text{fixed}}(t_l) + C_{\text{fuel}}(t_l) + C_{\text{carbon}}(t_l) \\
&= C_{\text{f}} \cdot \sum_{k \in K(t_l)} U_k(t_l) \\
&\quad + C_{\text{m}} \cdot \sum_{k \in K(t_l)} \sum_{b \in B_k(t_l)} \sum_{i \in N(t_l)} \sum_{j \in N(t_l) \& j \neq i} d_{ij} \cdot x_{ijk}^{b}(t_l)
\end{aligned}$$

TABLE 1. Notations of the EMDVRWC model.

Indices and Sets:	
i, j	index of the waste site, depot and disposal station
t_l	the scheduling point (l=0, 1, 2, …), t_0 is the initial scheduling point
k	index of the vehicle
b	index of the trip
$N(t_l)$	the set of depot, disposal station and the waste sites to be cleaned at t_l
$N_{new}(t_l)$	the set of waste sites with new collecting requirements at t_l
$N_{already}(t_l)$	the set of completed waste sites at t_l
$K(t_l)$	the set of available vehicles at t_l
$B_k(t_l)$	the set of trips of vehicle k at t_l
Parameters:	
Q	the maximum capacity of each vehicle
$L_k^b(t_l)$	the occupied capacity of vehicle k in the trip b at t_l
$Q_k^b(t_l)$	the remaining capacity of vehicle k in the trip b at t_l
$s_k(t_l)$	the last visited point of vehicle k before t_l (it may be a waste site, disposal station or depot)
z	the weight of the vehicle itself
q_i	the weight of waste at the waste site i
v	the average speed of each vehicle
$T_{\max}$	the maximum working hours of a driver
t_{k0}	the departure time of vehicle k
d_{ij}	the travel distance between point i and point j
l_{ij}^k	the load of vehicle k traveling from point i to point j
EC_{ij}^k	the carbon emission of vehicle k traveling from point i to point j
C_e	carbon tax
C_m	the unit fuel cost of a vehicle
C_f	the fixed cost of a vehicle
Decision variables:	
$y_{ik}^b(t_l)=1$	if vehicle k serves the site i in the trip b at t_l
$y_{ik}^b(t_l)=0$	else
$x_{ijk}^b(t_l)=1$	if vehicle k passes through the edge(i, j) in the trip b at t_l
$x_{ijk}^b(t_l)=0$	else
$U_k(t_l)=1$	if vehicle k is used at t_l
$U_k(t_l)=0$	else

$$+ C_e \cdot \sum_{k\in K(t_l)} \sum_{b\in B_k(t_l)} \sum_{i\in N(t_l)} \sum_{j\in N(t_l)\&j\neq i} EC_{ij}^k \cdot x_{ijk}^b(t_l) \quad (1)$$

$$\text{s.t.} \quad \forall k \in K(t_l), \sum_{b\in B_k(t_l)} \sum_{j\in N(t_l)} x_{s_k(t_l)jk}^b(t_l) = 1 \quad (2)$$

$$\forall i \in N(t_l)\backslash\{1,2\}, \sum_{k\in K(t_l)} \sum_{b\in B_k(t_l)} y_{ik}^b(t_l) = 1 \quad (3)$$

$$\forall k \in K(t_l),\ \forall b \in B_k(t_l),\ \forall j \in N(t_l)\backslash\{1,2\}, \sum_{i\in N(t_l)} x_{ijk}^b(t_l) = y_{jk}^b(t_l) \quad (4)$$

$$\forall k \in K(t_l),\ \forall b \in B_k(t_l),\ \forall i \in N(t_l)\backslash\{1,2\}, \sum_{j\in N(t_l)} x_{ijk}^b(t_l) = y_{ik}^b(t_l) \quad (5)$$

$$\forall k \in K(t_l),\ \forall j \in N(t_l)\backslash\{2\},\ l_{2j}^k(t_l) = 0 \quad (6)$$

$$\forall k \in K(t_l),\ \forall b \in B_k(t_l), \sum_{i\in N(t_l)} \sum_{j\in N(t_l)\&j\neq i} q_j \cdot x_{ijk}^b(t_l) \leqslant Q_k^b(t_l)$$

$$\forall k \in K(t_l), \quad (7)$$

$$(t_l - t_{k0}) + \sum_{b\in B_k(t_l)} \sum_{i\in N(t_l)} \sum_{j\in N(t_l)\&j\neq i} \frac{d_{ij}}{v} \cdot x_{ijk}^b(t_l) \leqslant T_{\max} \quad (8)$$

where Eq. (1) is the objective function, which minimizes the sum of the fixed cost ($C_{\text{fixed}}(t_l)$), the fuel cost ($C_{\text{fuel}}(t_l)$) and the carbon emission cost ($C_{\text{carbon}}(t_l)$) at

t_l. $EC_{ij}^k = FE \cdot FC_{ij}^k$, FE is the fuel emission parameter, and FC_{ij}^k is the fuel consumption of the vehicle k traveling from point i to point j. FC_{ij}^k can be obtained by Eq. (9) [47]:

$$FC_{ij}^k = \left[\alpha_{ij}(z + l_{ij}^k) + \beta v^2\right] d_{ij} \tag{9}$$

where α_{ij} and β are parameters related to road conditions and vehicle types. α_{ij} and β can be obtained by Eqs. (10) - (11).

$$\alpha_{ij} = a + g \sin\theta_{ij} + gC_r \cos\theta_{ij} \tag{10}$$

$$\beta = 0.5 C_d A \rho \tag{11}$$

where a is the acceleration of the vehicle, g is the acceleration of gravity, θ_{ij} is the road gradient from point i to point j, C_r is the coefficient of rolling resistance, C_d is the coefficient of traction, A is the frontal surface area of the vehicle and ρ is the density of air.

Constraint (2) guarantees that the starting point of vehicle k is its last visited site before t_l; Constraint (3) indicates that each waste site should be served only once by just one vehicle; Constraints (4) and (5) ensure that there is a vehicle driving from another site to it when each waste site is served and this vehicle departs from the waste site after completing the work; Constraint (6) indicates that each vehicle empties the waste it carries at the disposal station; Constraints (7) and (8) guarantee that the load and the working hours of each vehicle should not be greater than its maximum value, respectively.

IV. Q-LEARNING-BASED HYPERHEURISTIC PARTICLE SWARM OPTIMIZATION

With the aim of solving EMDVRWC, a predictive-rescheduling method based on QLHPSO is proposed. A dynamic response mechanism is introduced to quickly react to the occurrence of dynamic events. Eight low-level heuristics (LLHs) are designed, and a high-level strategy (HLS) is adopted to find the right LLH for the population with different states.

A. FRAMEWORK OF THE PREDICTIVE-RESCHEDULING METHOD

EMDVRWC is solved in a predictive-rescheduling manner, where the original routes can be adapted to the changing environment caused by dynamic events. The main framework of the predictive-rescheduling method is given in Fig. 1. First, at the initial time t_0, the optimal route is determined based on QLHPSO. Next, the vehicles collect the waste according to the initial optimal route until a dynamic event occurs. The time when the lth dynamic event occurs is regarded as the rescheduling point t_l. At t_l, a dynamic response mechanism is introduced to generate a heuristic initial population and the optimal route is rescheduled in the new environment by QLHPSO. Then the new schedule is executed until the next dynamic event triggers a QLHPSO-based rescheduling procedure. This process continues until all the waste collection tasks are completed. The flowchart of QLHPSOis shown in Fig. 2, which is mainly composed of initialization, objective evaluations, updating the personal best and the global best, selecting LLH by a $Q-$ learning-based high-level strategy (QHLS) and generating the new population by the selected LLH. The pseudocode of QLHPSO is shown in Algorithm 1.

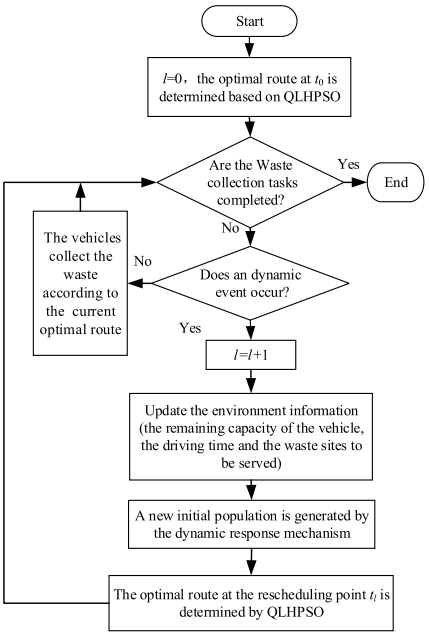


FIGURE 1. Framework of the predictive-rescheduling method for EMDVRWC.

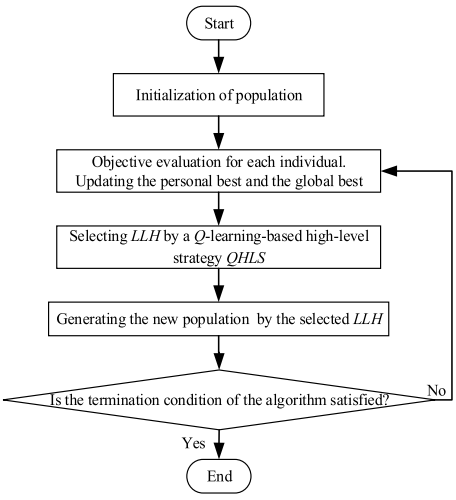


FIGURE 2. Flowchart of the proposed QLHPSO.

B. ENCODING AND DECODING OF INDIVIDUALS

To solve EMDVRWC, the integer encoding is used in the proposed algorithm QLHPSO. For a problem containing $n(t_l)$ points at t_l(1: depot, 2: disposal station, 3$\sim n(t_l)$: waste sites), each individual is encoded as a sequence of length $(n(t_l)-2)$, and each element takes a different integer from 3$\sim n(t_l)$. The

Algorithm 1 Q-learning-based hyperheuristic particle swarm optimization $QLHPSO(FE_{max}, pop, N)$

Input: FE_{max} – the number of objective evaluations, *pop* - population, N – population size
Output: x_{best} – the global best solution

```
FE ←0;
While FE ≤FE_max do
 for i= 1 to N do
   The objective is evaluated according to Eq. (1) for the ith particle of pop;
   FE ← FE+1;
   Updating the personal best of the ith particle;
   Updating the global best x_best;
 end for
 Selecting LLH by the Q−learning-based high-level strategy QHLS according to Algorithm 2;
 newpop← Applying the selected LLH to pop to generate a new population;
 pop ← newpop;
 end while
 Output x_best
```

greedy decoding is adopted, where individuals are decoded sequentially so that the vehicle utilization is maximized within the working hours.

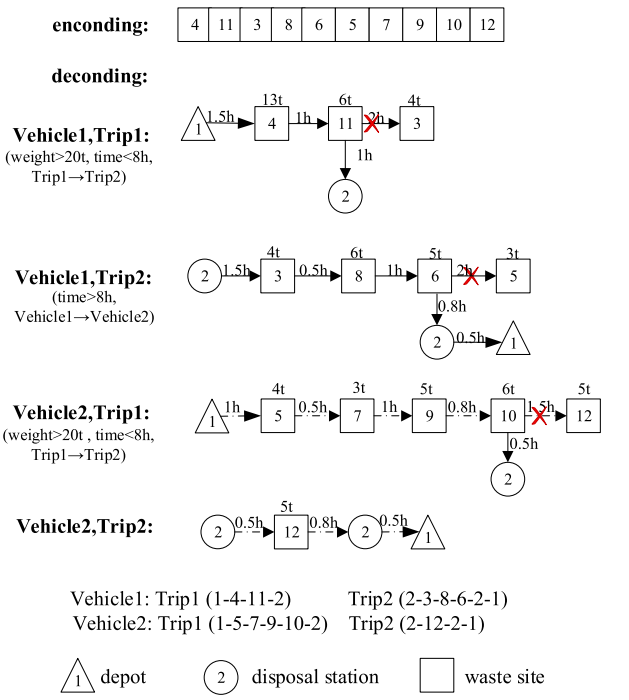


FIGURE 3. An illustration of the proposed encoding and decoding method.

Fig. 3 shows an encoding and decoding illustration with ten waste sites (numbered 3∼12). The waste weight of each site is placed above the site number, and the value on each arrow represents the time required to pass through the edge. Assume that the maximum capacity of each vehicle is 20T and the maximum working time is 8h. The method of decoding is detailed as follows: first, Vehicle 1 starts its first trip (Trip1) from the depot and serves the waste sites in the encoding sequentially. If Vehicle 1 continues to serve waste site 3 after finishing waste sites 4 and 11, its load will be 23T, exceeding the maximum capacity of the vehicle. Thus, Vehicle 1 returns to the disposal station for unloading after serving site 11.

Vehicle 1 then proceeds from the disposal station for its second trip (Trip2). Waste site 3 will be the first site to be served by Vehicle 1 in Trip 2, and the remaining waste sites are served in sequence. If Vehicle 1 continues to visit waste site 5 after completing waste site 6, the vehicle will work up to 8.5 hours, exceeding the maximum working hours of the driver. Therefore, after serving waste site 6, Vehicle 1 will go to the disposal station to unload and then return to the depot, where it ends its one day's journey. Vehicle 2 serves the remaining waste sites of the encoding in the same way. In this example, all the waste in the ten sites is collected by two vehicles. It can be seen from the above process that our decoding method can ensure feasibility of the obtained solutions since both the capacity and the time constraints are satisfied. Thus, we can handle all the constraints (Eqs. (2) - (8)) through our decoding method, without the need to design specific constraint handling methods. Moreover, the two subproblems of vehicle selection and vehicle routing can be handled simultaneously.

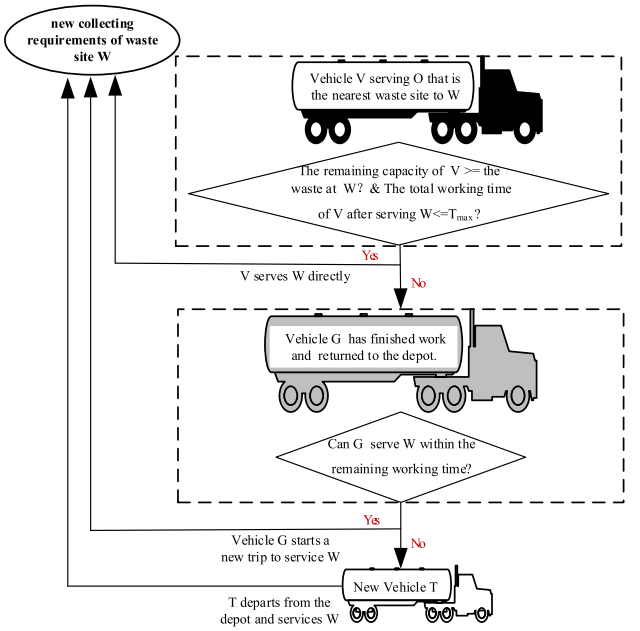
C. DYNAMIC RESPONSE MECHANISM

To provide a good search starting point for the algorithm, the dynamic characteristics of EMDVRWC are utilized. When a dynamic event occurs and the QLHPSO-based rescheduling method is triggered, a dynamic response mechanism (DRM) is introduced to generate the initial population, which includes the schedule repair, reuse of the historical information and random initialization.

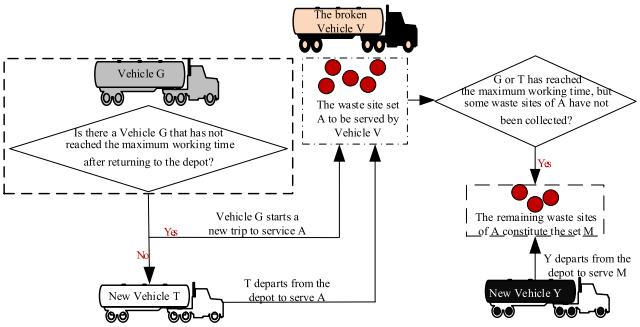
1) SCHEDULE REPAIR

The following strategies are designed to repair the original route plan by analyzing the characteristics of two dynamic events in EMDVRWC.

When new collecting requirements appear in the waste site W, each vehicle serves the original waste site according to the previous schedule. For the waste site W, the approach is shown in Fig. 4(a). First, the nearest original waste site O to W is determined. Then, if the remaining capacity of Vehicle V is enough to meet the collection requirements of W in the current trip, and the total working time of V after serving W does not exceed the maximum working time, Vehicle V will serve W. If Vehicle V does not meet the above conditions, judge whether Vehicle G can serve W within the remaining working time after vehicle G has finished work and returned to the depot. If it exists, Vehicle G will start a new trip to serve W. If it does not exist, a new Vehicle T will depart from the depot and serve W.



(a) Schedule repair for new collecting requirements of waste sites



(b) Schedule repair for vehicle breakdowns

FIGURE 4. Two schedule repairs based on characteristics of dynamic events.

When a vehicle V breaks down at work, the order of each working vehicle to serve the stations is kept unchanged. For the waste site set A that cannot be serviced due to the breakdown of V, take measures as shown in Fig. 4(b). First, judge whether there is a Vehicle G that has not reached the maximum working time after returning to the depot (according to Part B of Section **IV**, no more than one car meets this condition). If it does not exist, a new Vehicle T will directly depart from the depot to serve A. If it exists, Vehicle G will start a new trip after returning to the depot to serve A. Thereafter, if Vehicle G or T has reached the maximum working time but some waste sites of A have not been completely collected, a new Vehicle Y will depart from the depot to serve the remaining waste site set M.

2) REUSE OF THE HISTORICAL INFORMATION

At each scheduling point, the optimal routes generated at the previous scheduling point can be reused as historical information. At the current point, the service order of all available vehicles for original waste sites can be referred to the historical solution.

3) RANDOM INITIALIZATION

To introduce diversity, random individuals are introduced in the initial population where waste sites and the order of service are randomly generated.

In the initial population, repair solutions and their variants account for 30%, the historical solution and its variants accounted for 20% and random individuals account for 50%.

D. FRAMEWORK OF THE HYPERHEURISTIC ALGORITHM

Framework of the hyperheuristic algorithm is shown in Fig. 5, which is mainly composed of two levels: the problem domain and the HLS [48]. The problem domain is composed of a series of LLHs, problem definitions, evaluation functions and initial solutions, etc. At the HLS level, a new heuristic algorithm is constructed or an appropriate LLH is selected by manipulating and managing the LLH library with the help of domain knowledge. The two levels are independent of each other. Once information on the problem domain is modified, the hyperheuristic algorithm can be quickly adapted to the new problems, which is generic and efficient.

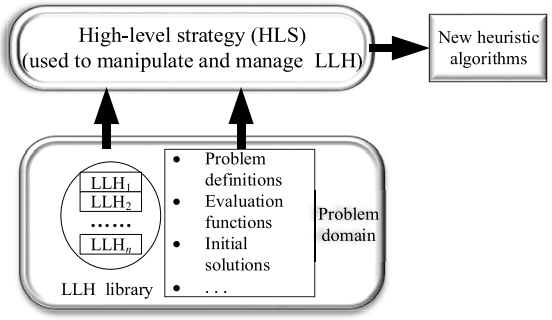


FIGURE 5. Framework of the hyperheuristic algorithm.

E. LOW-LEVEL HEURISTICS

QLHPSO includes learning operators and local search operators, the combinations of which are regarded as different LLHs.

1) FOUR LEARNING OPERATORS

(a) Greedy learning (GL). At the early stage of evolution, the population is in a rapid convergence stage and needs to approach the extremum quickly. For this, the GL operator is designed according to the characteristics of EMDVRWC. First, considering that EMDVRWC is an asymmetric problem [49], the cost on the same path with opposite directions might be different from each other. Thus, the reverse mutation (arranging the routes between two points of an individual in a reverse order) is adopted to imitate the 'inertia' part of the PSO velocity update formula,

where the unreversed parts of the individual are preserved as inertia. Then, the greedy crossover operator is applied to reflect the 'self-learning ability' and 'social learning ability' in the PSO velocity update formula. Since the crossover operator allows the information interaction between two discrete individuals, it can make the individual learn from the personal best and the global best. Inspired by the idea of greedy search, the information on vehicle capacity and distance is introduced into the crossover operator so that the transportation cost can be reduced. The proposed GL operator not only makes PSO suitable for solving discrete problems like EMDVRWC, but also accelerates the convergence of the algorithm.

(b) Multi-operator Learning(ML). When the population has evolved for a certain number of generations adopting the GL operator, most particles concentrate in the same region rapidly due to the use of heuristic information. In this case, the searching range of the population is very limited, leading to the premature convergence. Thus, at this stage of evolution, learning strategies should be adjusted, which can maintain population diversity while producing new individuals. With this in mind, a ML operator is designed. The 'inertia' part of the PSO velocity update formula is replaced by the multi-mutation operator, where one of the three methods of swap mutation, reverse mutation and insertion mutation is selected randomly to perform mutation on each individual. Meanwhile, the multi-crossover operator is employed as the way to learn from the personal best and the global best, where one of the three operators of the partially mapped crossover, sequential crossover and cyclic crossover is randomly selected. Compared with the GL operator, the ML operator provides a richer way to generate new individuals, which improves the population diversity effectively.

(c) Exploring learning(EL). It has been proved in Liang et al. [50] that it is more conducive to global search by making each dimension of an individual learn randomly from the personal best of its own or other individuals. Inspired by this, the EL operator is proposed, which still adopts the multi-mutation operator to replace 'inertia' and the multi-crossover operator to learn from the elite individual, like those used ML. The difference in EL and ML is that each dimension of the current individual can learn stochastically and independently from the personal best of any other individual in the population, as long as it has a better fitness than the current one. The EL operator enriches the sources of information interaction among individuals, which diversifies the searching directions of population while guaranteeing the quality of individuals.

(d) Exploiting learning(EXL). With the aim of improving the local search ability of the algorithm, the EXL operator is proposed, where each individual only learns from the personal best. The multi-mutation operator is used to replace 'inertia', and the partially mapped crossover is adopted to realize the interaction between the individual and the personal best. EXL makes full use of the useful information of the individual itself, which is beneficial to mine a better solution around the personal best.

2) TWO ENHANCED LOCAL SEARCH OPERATORS

Enhanced local search #1 (ELS1): In order to conduct a fine search around the decoded individual, an enhanced local search is designed based on the classical 2-opt operator [51], which helps open up the intersection possibly existing in the longest trip. The implementation of ELS1 is illustrated in Fig. 6. First, the individual X is decoded according to Part B of Section **IV**, where two vehicles are needed and each vehicle has two trips. Then, for each vehicle, the trip that contains the largest number of points (including waste sites, depot and disposal station) is found out. Next, the 2-opt operator is conducted on such trips (Trip 2 of Vehicle 1 and Trip 1 of Vehicle 2) to find new trips with lower cost. Finally, the optimized trips (2-6-8-11-4-9-2-1), (1-13-16-10-12-14-2) and the unselected trips (1-3-5-7-2), (2-18-15-17-2-1) are recombined to form a new individual X_{new}. Note that the depot (marked as 1) and the disposal station (marked as 2) should be removed before recombining. If $X_{new}$is better than Xin terms of the total cost, then X is replaced by X_{new}. The above procedure is performed Ytimes around each decoded individual.

Enhanced local search #2 (ELS2): The procedure of ELS2 is consistent with ELS1, while the difference is that ELS2 employs the point interpolation [52] instead of the 2-opt operator on the trips with the most points. ELS2 is beneficial to find a better solution in the case of no crossing of the route.

3) EIGHT LOW-LEVEL HEURISTICS

Combining the four learning operators with the two enhanced local search operators, a total of eight LLHs are defined, as shown in Table 2.

F. A Q-LEARNING-BASED HIGH-LEVEL STRATEGY

A $Q-$ learning-based high-level strategy (QHLS) is designed. $Q-$ learning is a value-based reinforcement learning algorithm which aims to produce a $Q-$table (i.e., a state-action pair table) as shown in Table 3.$Q-$value in the table is the maximum expected future reward that can be obtained when each action is taken in each state. Q-learning selects the best action for a given state based on the Q-table. The implementation of QHLS is shown in Algorithm 2. If the generation $t = 1$, all Q-values in the Q-table are initialized to be 0, where state is the population state set and LLH_pool is the LLH set (lines 1-2). Otherwise, the Q-table (lines 3-6) is updated based on the reward obtained by performing the action in the previous population state. Then, the current population state is perceived (line 7). Finally, the selection probability of each LLH in the current population state is calculated (line 8) and the one to be employed is determined through roulette selection (line 9). The state, the selection mechanism of the LLH, the reward function and update of the Q-table in QHLS are respectively introduced as follows.

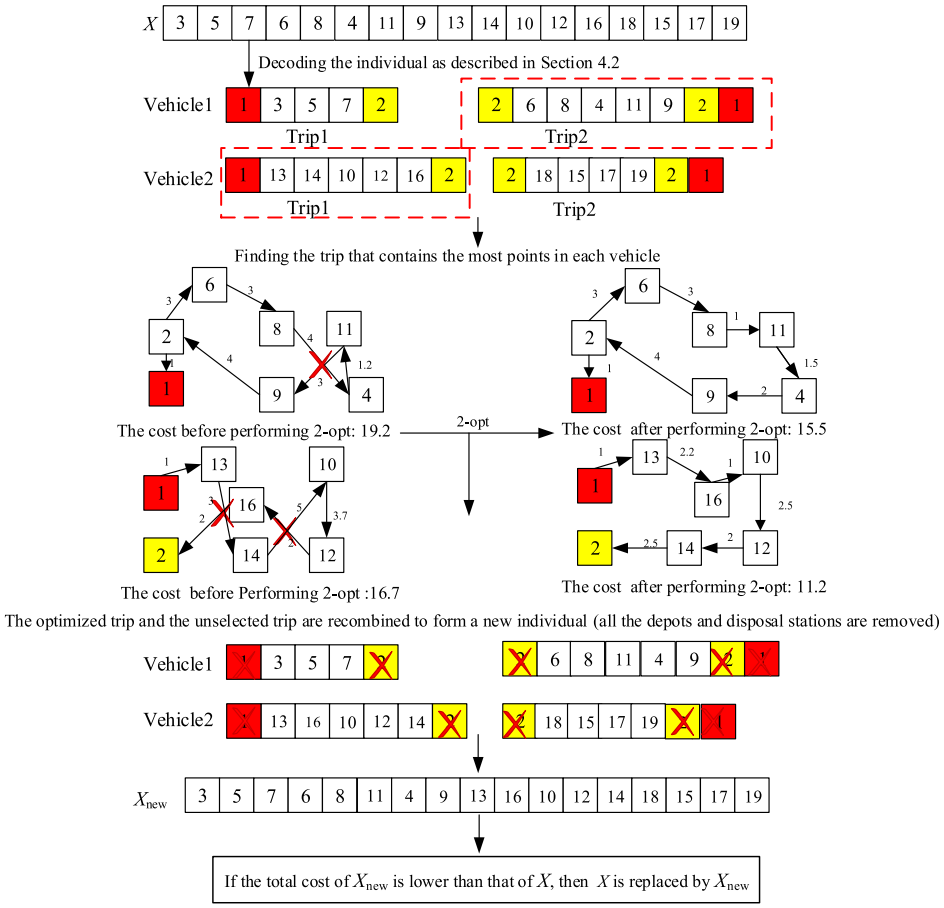


FIGURE 6. Implementation of the enhanced local search ELS1.

TABLE 2. Eight LLHs.

LLH_1	LLH_2	LLH_3	LLH_4	LLH_5	LLH_6	LLH_7	LLH_8
$GL+ELS_1$	$GL+ELS_2$	$ML+ELS_1$	$ML+ELS_2$	$EL+ELS_1$	$EL+ELS_2$	$ELX+ELS_1$	$ELX+ELS_2$

TABLE 3. *Q*-table.

state	action					
	LLH_1	LLH_2	LLH_3	LLH_4	.....	LLH_8
S_1	$Q(S_1, LLH_1)$	$Q(S_1, LLH_2)$	$Q(S_1, LLH_3)$	$Q(S_1, LLH_4)$	.....	$Q(S_1, LLH_8)$
S_2	$Q(S_2, LLH_1)$	$Q(S_2, LLH_2)$	$Q(S_2, LLH_3)$	$Q(S_2, LLH_4)$	.....	$Q(S_2, LLH_8)$
S_3	$Q(S_3, LLH_1)$	$Q(S_3, LLH_2)$	$Q(S_3, LLH_3)$	$Q(S_3, LLH_4)$	.....	$Q(S_3, LLH_8)$
S_4	$Q(S_4, LLH_1)$	$Q(S_4, LLH_2)$	$Q(S_4, LLH_3)$	$Q(S_4, LLH_4)$	.....	$Q(S_4, LLH_8)$

1) DEFINITION OF THE STATE

QHLS selects the LLH for the population according to the population state. Therefore, the division of the population state is important for the selection of LLHs. We define the population state considering the convergence and diversity, both of which are the key performance indicators in evolutionary algorithms. Convergence is defined as the increment of the optimal fitness value between two consecutive moments. Assume $x_{\text{best}}(t)$ and $x_{\text{best}}(t-1)$ are the optimal solutions in the tth and $(t-1)$th generation, respectively. Let $\Delta f(x_{\text{best}}) = f(x_{\text{best}}(t)) - f(x_{\text{best}}(t-1))$. Obviously, $\Delta f(x_{\text{best}}) > 0$ indicates the improvement of convergence,

Algorithm 2 The Q-learning-based high-level strategy $QHLS$ (t, $S(t\text{-}1)$, $LLH(t\text{-}1)$, $x_{\text{best}}(t)$, Q)

Input: t- current generation, $S(t-1)$−state at the previous generation, $LLH(t-1)$−the LLH used in $S(t-1)$, $x_{\text{best}}(t)$−the current best solution at t, Q−Qtable

Output: $LLH(t)$−the LLH selected in the current state $S(t)$, Q-the updated Q table

if $t$equals to 1
  $Q(S, LLH_i) \leftarrow 0,\ \forall S \in State, \forall LLH_i \in LLH\_pool$; // Initialize the Q−table
else
  According to Eq. (17), the reward $r(t-1)$ for $LLH(t-1)$ under $S(t-1)$ is calculated;
  Update $Q(S(t-1), LLH(t-1))$ according to Eq. (18);
end if
The current state $S(t)$ of the population is perceived based on the current optimal solution $x_{\text{best}}(t)$;
According to Eq. (16), the selection probability $P(S(t), LLH_i)$ of each LLH in $S(t)$ is calculated;
$LLH(t) \leftarrow Roulette(P(S(t), LLH_i))$; //The low-level heuristic $LLH(t)$ in $S(t)$ is determined by roulette selection
  Output $LLH(t)$, Q.

while $\Delta f(x_{\text{best}}) = 0$ suggests the stagnation of the population.

The population diversity is defined as the similarity of population, where a higher similarity indicates a worse diversity. Considering the characteristics of the solution to EMDVRWC, the similarity between two routes is defined based on the number of the same edges, as shown in Fig. 7.

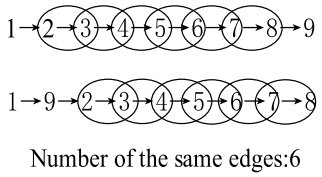


FIGURE 7. Definition of the similarity between two routes.

First, an edge connection matrix $G = (x_{ij})_{n_1 \times n_1}$ is established for an individual of length n_1(Eq. (12)). If the individual contains the edge from point i to point j, then $x_{ij} = 1$; otherwise, $x_{ij} = 0$. Second, for any two edge connection matrices G_1 and G_2, the set $D(G_1, G_2)$ which includes the same edges in them is obtained by Eq. (13), and $similar\ (G_1, G_2) \in [0, 1]$ is calculated according to Eq. (14). Finally, the population similarity $similar_{pop} \in [0, 1]$ is defined as Eq. (15), where PS is the population size. The similarity between each individual and the global optimal individual x_{best} is calculated, then the mean value is obtained to describe the aggregation degree of the population.

$$G = (x_{ij})_{n_1 \times n_1} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & \ldots & 0 \end{bmatrix}_{n_1 \times n_1} \tag{12}$$

$$D(G_1, G_2) = \left\{ x_{ij}^{G_1} | x_{ij}^{G_1} = 1 \ \&\ x_{ij}^{G_2} = 1,\ \forall i, j \in \{1, 2, \cdots, n_1\} \right\} \tag{13}$$

$$similar(G_1, G_2) = \frac{|D(G_1, G_2)|}{n_1 - 1} \tag{14}$$

$$similar_{pop} = \frac{\sum_{k=1}^{PS} similar(G_k, x_{\text{best}})}{PS} \tag{15}$$

According to the population convergence and diversity, four states are divided as listed in Table 4.

2) SELECTION MECHANISM OF LLHs

Assume that the current population state is S(t). The selection probability of each candidate LLH_i ($i = 1,2, \ldots, num$) is calculated from the $Q-$ value of the state-action pair in Q−table, as shown in Eq. (16). θ is set as 2 [19] and the roulette selection is used to determine the LLH to be employed in $S(t)$.

$$P(S(t), LLH_i) = \frac{\exp(\theta Q(S(t), LLH_i) / \max_{i=1,2,\ldots,num} Q(S(t), LLH_i))}{\sum_{i=1}^{num} \exp(\theta Q(S(t), LLH_i) / \max_{i=1,2,\ldots,num} Q(S(t), LLH_i))} \tag{16}$$

3) REWARD FUNCTION

Reward function is used to confirm the search performance of the most effective LLH in each state. The direct return of a search operator applied to a certain state depends on the degree to which the operator improves fitness of the optimal solution and diversity of the population. Eq. (17) is designed as the reward function. When the generated new solution is superior to the previous optimal solution in terms of the fitness value, the reward value is calculated according to the improvement degree of fitness. When there is no difference in fitness, but the population diversity is better than that of the previous generation, the reward value is calculated according to the improvement degree of diversity. Otherwise, the reward

TABLE 4. Division of the state.

S_1	S_2	S_3	S_4
$0 \le similar_{pop} < 0.6$	$0 \le similar_{pop} < 0.6$	$0.6 \le similar_{pop} \le 1$	$0.6 \le similar_{pop} \le 1$
$\Delta f\left(x_{\text{best}}\right) > 0$	$\Delta f\left(x_{\text{best}}\right) = 0$	$\Delta f\left(x_{\text{best}}\right) > 0$	$\Delta f\left(x_{\text{best}}\right) = 0$

is 0.

$$r(t-1) = \begin{cases} \dfrac{f\left(x_{\text{best}}(t-1)\right)}{f\left(x_{\text{best}}(t-2)\right)} \; if \; f\left(x_{\text{best}}(t-1)\right) > f\left(x_{\text{best}}(t-2)\right) \\ \dfrac{similar_{pop}(t-2)}{similar_{pop}(t-1)} if \; f\left(x_{\text{best}}(t-1)\right) = f\left(x_{\text{best}}(t-2)\right) \\ \text{and } similar_{pop}(t-1) < similar_{pop}(t-2) \\ 0 \; otherwise \end{cases} \tag{17}$$

4) UPDATE OF THE Q-TABLE

The $Q-$value of $Q(S(t-1), LLH(t-1))$ reflects the search ability of the low-level heuristic $LLH(t-1)$in state $S(t-1)$, where the initial $Q-$ value of $Q-$ table is 0. In iterations, $Q-$table is updated as Eq. (18).

$$\begin{aligned} &Q(S(t-1), LLH(t-1)) \\ &= (1-\alpha)\, Q(S(t-1), LLH(t-1)) \\ &+ \alpha \left[r(t-1) + \gamma \max_{LLH_i \in \{LLH_1, LLH_2, \ldots, LLH_{num}\}} Q(S(t), LLH_i) \right] \end{aligned} \tag{18}$$

where $\max_{LLH_i \in \{LLH_1, LLH_2, \ldots, LLH_{num}\}} Q(S(t), LLH_i)$ is the maximum $Q-$value of all the actions in the population state $S(t)$, which is obtained after executing $LLH(t-1)$ in state $S(t-1)$. $0 \leqslant \alpha \leqslant 1$ is the learning rate and $0 \leqslant \gamma \leqslant 1$ is the discount rate indicating the influence of future rewards on the current state. $\gamma = 0$ suggests that $Q-$ learning only pays attention to immediate rewards.

G. COMPLEXITY ANALYSIS OF THE ALGORITHM

At the scheduling point t_l, assume the population size is N and the individual dimension is $D(t_l)$, then the time complexity of the standard PSO is $O(N^* D(t_l))$. For QLHPSO, the enhanced local search in LLHs needs to search Y times around the neighborhood of each decoded individual, so the time complexity of QLHPSO is $O(N^* D(t_l)^* Y)$. Although the time complexity of QLHPSO is Y times than that of PSO, the experimental results in Section V show that its search accuracy is significantly improved. Since all the original individuals in QLHPSO are replaced by the updated individuals and no additional space is occupied, QLHPSO has the same space complexity as standard PSO, which is $O(N^* D(t_l))$.

V. EXPERIMENTAL STUDIES

Two groups of experiments on ten EMDVRWC instances are performed to verify the effectiveness of the novel strategies and the overall performance of the proposed algorithm, respectively. All experiments are implemented in MATLAB R2017b running on a personal computer with Intel (R) Core (TM) i5-5500U and 16 GB running memory.

A. INSTANCE GENERATION AND PARAMETER SETTINGS

We make an investigation on the Green Ring Company of Jiangbei New District in Nanjing of China and obtain a real-world EMDVRWC instance on waste collection. This instance includes the coordinates of 54 waste sites distributed in 54 residential areas, one depot and one disposal station in Jiangbei New Area. The daily waste amount of each residential area is also covered. Besides, to measure the scalability of the proposed algorithm to problems with different sizes, nine synthetic instances ranging from 17 to 201 customer points are selected from CVRP datasets, including set A, set B, set E and set CMT [53]. In the original CVRP datasets, both the customer demands and the distances between customer points are large, which do not conform to the actual situation of waste collection in residential areas. Thus, the coordinates and demands in the nine synthetic instances need to be adjusted according to the data acquired from the investigation on the real company. All the coordinates are decreased by the same scale, which not only reduces the distances between customers, but also maintains their relative positions. The demands of customers are processed in the same way. The depot and the disposal station are respectively marked as Point 1 and Point 2, the coordinates of which are generated randomly in a certain range.

To further assess the algorithm's robustness and adaptability to various dynamic scenarios, two kinds of dynamic events (new collecting requirements of waste sites and vehicle breakdowns) are introduced into the real-world and synthetic instances. At the initial time of each instance, there are n points, including $(n-2)$ waste sites, a depot and a disposal station. Then it is assumed that two kinds of dynamic events occur one by one following a Poisson distribution, i.e., the time interval between the occurrence of two events is assumed to follow an exponential distribution. In this way, the occurring time, the occurring order and the number of dynamic events are all different on each instance. Thus, different instances have various dynamic scenarios. The x and y coordinates of the waste sites with new collecting requirements are randomly generated

within [10], [50], and the waste amount of such waste sites is randomly generated within [0.5, 2]. Through the above procedure, a total of ten EMDVRWC instances are generated and each is named JTn−DTc (n is the number of points at the initial time and c is the number of dynamic events). Particularly, the real-world instance is named JT56-DT3. Parameter settings in the EMDVRWC model are shown in Table 5 [54].

The parameter values in the Q-learning and PSO components of the proposed algorithm QLHPSO were set based on the literature. The population size N is set to 200 and the number of local search iterations Y is set to 10, which are commonly-used in the standard PSO [50] and the 2-opt operator [51]. The learning rate in Eq. (18) is set as $\alpha = 1 - (0.9 * NFE/NFE_{max})$ [48], where NFE and NFE_{max} are the current and the maximum number of objective evaluations, respectively, and the discount rate γ is set as 0.8 [48]. In the future, we can perform an orthogonal experiment with specific levels to further determine the best setting of each parameter on each instance.

B. PROCEDURE OF THE EXPERIMENTS

Assume that a EMDVRWC instance contains one initial scheduling point and $(L-1)$ rescheduling points (different instances might take different values of L). Nine algorithms are adopted to solve the problem at each point (the proposed algorithm, four comparison algorithms in validations of the strategies in Part C of Section **V**, and four comparison algorithms in validations of the overall performance of the algorithm in Part D of Section **V**). Since dynamic problems have multiple scheduling points, the performance evaluation of each algorithm over all scheduling points differs from static scheduling. The experimental procedure of dynamic scheduling on EMDVRWC is detailed as follows:

Step1: At each scheduling point (including the initial scheduling point and the subsequent rescheduling points), each of the nine algorithms performs 20 independent runs. Each algorithm terminates until the number of objective evaluations reaches the maximum value $NFE\max = 50000$, i.e., different algorithms have the same number of objective evaluations in one run. Population size of all the algorithms is set as 200.

Step2: At each scheduling point, the optimal solution with the minimum objective value is determined from all the ''optimal solutions'' obtained by nine algorithms in each of the 20 runs, which is employed as the new waste collection schedule after the current scheduling point. This manner ensures that nine algorithms are compared at each scheduling point in the same environment.

Step3: The waste collection tasks continue to be executed based on the optimal schedule.

Step4: If the tasks have not been completed and new dynamic events occur, then move to the next scheduling point and go to Step 1. Otherwise, Step 5 is conducted after all the tasks are finished.

Step5: Compare the overall performance of nine algorithms across different scheduling points and runs. As shown in Fig. 8, for the j^{th} (j=1, 2, …,20) run of the k^{th} (k=1, 2, …,9) algorithm, the optimal objective values (Eq. (1)) obtained at all scheduling points (t_0, t_1,…, t_l,…, t_L) are averaged to get $mean_j^k$. The 20 mean values obtained from 20 runs form the vector Vec_k. The optimal and average value of the k^{th} algorithm in Vec_k are denoted as Best_k and mean_k, which are regarded as the overall optimal value and average value of the k^{th} algorithm, respectively. Finally, the Wilcoxon signed-rank tests with the significance level of 0.05 are adopted to perform statistical tests on the vector of the proposed algorithm and those of the other eight algorithms. Comparison results of the nine algorithms on all the ten EMDVRWC instances are presented in Table 6.

C. VALIDATING THE EFFECTIVENESS OF THE NOVEL STRATEGIES

This section performs an ablation study, which validates the effectiveness of the three novel strategies introduced in the proposed algorithm, including the dynamic response mechanism DRM, definition of the population state and the Q-learning-based high-level strategy QHLS.

1) VALIDATING THE EFFECTIVENESS OF DRM

To verify the effectiveness of DRM in Part C of Section **IV**, the DRM in the proposed algorithm QLHPSO is replaced with random initialization, which results in a comparison algorithm QLHPSO_RI. The comparison results between QLHPSO and QLHPSO_RI are listed in Table 6.

It can be seen from Table 6 that QLHPSO_RI obtains the same ''Best'' and ''mean'' values as QLHPSO only on two small-scale instances JT17-DT3 and JT33-DT2. On other eight instances, including the real-world instance JT56-DT3, the ''Best'' and ''mean'' values produced by QLHPSO are better than those of QLHPSO_RI. Meanwhile, the Wilcoxon rank sum test results show that QLHPSO is significantly better than QLHPSO_RI on seven instances, including the real-world instance JT56-DT3, and there is no significant difference between them only on three small-scale instances. These results indicate that the proposed DRM-based population initialization behaves better, both on the real instance and the synthetic ones. The reason is that when rescheduling, a high-quality initial population is regenerated by incorporating 20% historical solutions and 30% schedule repair solutions. It utilizes the features of the dynamic events and elite information of the pervious scheduling point, which reduces the blind search at the initial stage. In addition, 50% random individuals are also introduced into the initial population which guarantees the population diversity.

TABLE 5. Parameter settings of the EMDVRWC model.

notation-meaning	value	notation-meaning	value
z-the weight of the vehicle itself	1 (t)	g-the gravitational acceleration	9.81 (m/s^2)
Q-the capacity of each vehicle	10 (t)	θ-the slope of the road	0
T_{max}-the maximum working hours of a driver	8 (h)	FE-the fuel emission parameter	2.621×10^{-6} (t/L)
v-the average speed of each vehicle	30 (km /h)	C_d-the traction coefficient	0.7
C_e-the carbon tax	5 (CNY/t)	C_r-the rolling resistance coefficient	0.3
C_m-the unit fuel cost of a vehicle	2 (CNY/km)	A-the front surface area of the vehicle	5 (m^2)
C_f-the fixed cost of a vehicle	500 (CNY)	ρ-the air density	1.204 (kg/m^3)
a-the acceleration of a vehicle	0		

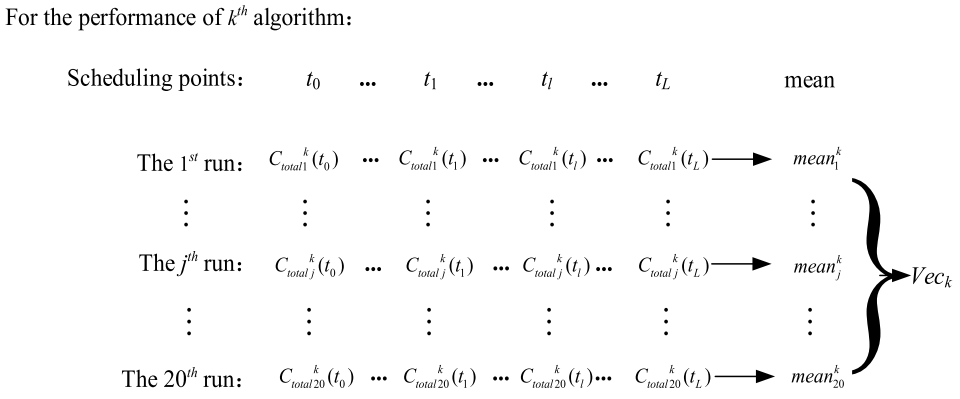


FIGURE 8. The overall performance of the kth algorithm in an EMDVRWC instance.

2) VALIDATING THE DEFINITION OF POPULATION STATES

In QLHPSO, the population state is defined based on convergence and diversity, as described in Part F of Section **IV**. With the aim of verifying the proposed state division method, it is replaced by the method in Gölcük and Ozsoydan [48], which divides the population state based on fitness value. The resulting algorithm is named QLHPSO_fit and it is compared with QLHPSO. The comparison results are shown in Table 6.

Table 6 shows that QLHPSO_fit obtains the same "Best" values as QLHPSO on two small-scale instances JT17-DT3 and JT33-DT2, and produces the same "mean" value only on JT17-DT3. On other eight instances, including the real-world instance JT56-DT3, the "Best" and "mean" values searched by QLHPSO are better than those of QLHPSO_fit. Meanwhile, the Wilcoxon rank sum test results show that QLHPSO is significantly better than QLHPSO_fit on seven instances, including the real-world instance JT56-DT3, and there is no significant difference between them only on three small-scale instances. These results indicate that our way to divide the state is more effective, both on the real instance and the synthetic ones. Compared with the dividing method based on fitness value, our method measures the population state from two aspects of convergence and diversity, both of which are critical to the accuracy of the final solution. Thus, it can reflect the population status more comprehensively, providing a good foundation for the subsequent use of QHLS.

3) VALIDATING THE EFFECTIVENESS OF QHLS

In order to analyze the influence of QHLS on the algorithm performance, QHLS in QLHPSO is respectively replaced with the intelligent selection strategy of Moodi et al. [55] and the adaptive strategy of Sabar et al. [56], obtaining two algorithms named SSLPSO and APSO. The intelligent selection strategy determines the frequency and priority of using each learning strategy according to the number of changes in personal best and global best at each stage. The adaptive strategy encodes a set of configurations into the individual, which contains parameters, the serial number of each search operator, and the order in which each operator is called. Then, individuals can adaptively provide the suitable search method for the population by using different configurations during the search process. The proposed algorithm QLHPSO is compared with SSLPSO and APSO, the results of which are shown in Table 6.

As shown in Table 6, QLHPSO gets the same "Best" value as SSLPSO on JT17-DT3, JT81-DT3 and JT201-DT4, and obtains a worse value than SSLPSO on JT65-DT4. On the remaining six instances, including the real-world instance JT56-DT3, the "Best" value of QLHPSO is better than that

TABLE 6. Average performance values and the sign of statistical test results of nine algorithms across 20 different runs on the nine synthetic instances and one real-world instance (The best value is in bold. The sign of '+/-/=' in the comparison algorithm B indicates that the proposed algorithm QLHPSO is significantly better than B, significantly worse than B, or there is no significant difference between QLHPSO and B. The '+/=/-' column indicates the number of instances which take '+', '=' or '-').

Instance	JT17-DT3		JT23-DT2		JT33-DT2		JT56-DT3	
	Best	mean	Best	mean	Best	mean	Best	mean
QLHPSO	**1300**	**1300**	**2370**	2400	**2500**	**2520**	**5090**	**5290**
QLHPSO_RI	**1300**	**1300=**	2380	2400=	**2500**	**2520=**	5520	5620+
QLHPSO_fit	**1300**	**1300=**	2380	2400=	**2500**	2530=	5310	5390+
SSLPSO	**1300**	**1300=**	2380	2400=	2510	2530+	5220	5350+
APSO	**1300**	**1300=**	2390	2430+	2520	2590+	5370	5590+
GA_TS	**1300**	**1300=**	**2370**	**2380-**	2510	2530=	5640	5720+
SAEA	**1300**	**1300=**	2410	2440+	2530	2570+	5670	5740+
PSO	**1300**	**1330+**	2810	2890+	3020	3110+	6700	7030+
2MPSO	**1300**	**1300=**	2480	2550+	2550	2610+	5730	5780+
Instance	JT65-DT4		JT79-DT4		JT81-DT3		JT102-DT4	
	Best	mean	Best	mean	Best	mean	Best	mean
QLHPSO	4310	**4350**	**5350**	**5430**	**6990**	**7110**	**10600**	**10900**
QLHPSO_RI	4320	4400+	5530	5710+	7340	7690+	11500	11700+
QLHPSO_fit	4330	4670+	5440	5570+	7100	7290+	11000	11400+
SSLPSO	**4280**	4360=	5380	5500+	**6990**	7120=	10800	11000+
APSO	4400	4670+	5520	6220+	7280	8110+	11300	12000+
GA_TS	4390	4640+	5970	6370+	8300	8680+	12100	12500+
SAEA	4360	4480+	5650	5820+	7370	7620+	11000	11600+
PSO	6220	6390+	8420	8620+	9150	9420+	15700	16000+
2MPSO	4630	4770+	5960	6030+	7740	7910+	11500	11600+
Instance	JT152-DT4		JT201-DT4		+/=/-			
	Best	mean	Best	mean				
QLHPSO	**14700**	**15200**	**16900**	**17500**				
QLHPSO_RI	15400	16200+	18400	19200+	7/3/0			
QLHPSO_fit	15000	15700+	17600	18000+	7/3/0			
SSLPSO	15000	15300+	**16900**	17600=	5/5/0			
APSO	15200	16900+	17600	19700+	9/1/0			
GA_TS	19600	20200+	25800	26500+	8/1/1			
SAEA	15600	16100+	18100	18700+	9/1/0			
PSO	22700	23200+	27900	28500	10/0/0			
2MPSO	15300	15500+	17200	17600	8/2/0			

of SSLPSO. In addition, the "mean" value of QLHPSO is better than SSLPSO on all the instances except JT17-DT3. The "Best" and "mean" values obtained by APSO are the same as QLHPSO only on JT17-DT3, while worse than QLHPSO on all the other instances, including the real-world instance JT56-DT3. The Wilcoxon rank sum test results show that QLHPSO is significantly better than SSLPSO and APSO on five and nine instances, respectively, both of which include the real-world instance JT56-DT3.

The above results indicate that QHLS is able to improve the search accuracy of the algorithm, both on the real instance and the synthetic ones. The reason for the success of QHLS is that it can establish a mapping between the current population state and the search operators (i.e., LLHs). At a state, a large $Q-$value indicates that the corresponding search operator has a high probability to be chosen for generating high-quality offsprings. QHLS helps the algorithm learn to autonomously choose the appropriate evolutionary mode in different population states during iterations of trial and error, which takes full advantage of each LLH and enhances the exploitation ability of the algorithm significantly. In addition, owing to its practicability in self-adapting to any state without manual adjustment, QHLS also improves the robustness of the algorithm to various dynamic environments.

D. VALIDATING THE PERFORMANCE OF THE PROPOSED ALGORITHM

In order to evaluate the overall performance of the proposed algorithm QLHPSO in solving EMDVRWC, it is compared with four recent meta-heuristic algorithms originally designed for the classical dynamic vehicle routing problem (DVRP). The four algorithms are replicated and applied

to EMDVRWC, which include GA_TS [57], SAEA [56], PSO [58] and 2MPSO [59]. GA_TS is a hybrid algorithm, in which the solution generated by genetic algorithm is used as the initial point of tabu search, and then the near global optimal solution is further obtained. GA_TS uses the same encoding and decoding method as our QLHPSO (see Part B of Section **IV**). In SAEA, each individual is encoded as an array, which has three parts: the path segment of each vehicle, the parameter values involved in the operator and the operator. SAEA adaptively evolves the configuration (parameter values, the number of search operators and the order to use search operators) of evolutionary algorithms. PSO employs real encoding, which is mapped to an integer sequence and decoded into a multi-trip route in the same way as our QLHPSO. 2MPSO includes two stages. The first stage assigns the nearest route to each of the k vehicles by the greedy method. In the second stage, particle swarm optimization starts searching from the results obtained by the first stage, where individuals are converted to integer sequences and decoded by the same method as our QLHPSO. Population size of all the four comparison algorithms is set as 200. Other parameters follow their original literature. The comparison results of QLHPSO with the above four algorithms are shown in Table 6.

As shown in Table 6, the ''Best'' and ''Mean'' values found by QLHPSO are better than the four comparison algorithms on all the medium to large-scale instances (JT33-DT2 to JT201-DT4), including the real-world instance JT56-DT3. The Wilcoxon rank sum test results show that QLHPSO is significantly better than GA_TS, SAEA, PSO and 2MPSO on eight, nine, ten and eight instances, respectively, all of which include the real-world instance JT56-DT3. The above results show that QLHPSO has better overall performance than the comparison algorithms when solving EMDVRWC. It can also be found that QLHPSO behaves better not only in scalability to the problem size, but also in adaptability to the environmental changes. Once a dynamic event occurs, DRM provides a good starting point for the search of QLHPSO by incorporating heuristic information, which helps QLHPSO react to the changes quickly. In the framework of Q-learning based hyerheuristic, eight LLHs are designed to enrich the behavior patterns of particles, helping the population jump out of the local optimum. Meanwhile, QHLS provides a well-directed search in the dynamically changing environment for it aims to determine the most suitable LLH for different population states, which are measured by convergence and diversity. All the above strategies are beneficial to rescheduling a near global optimal vehicle routing solution with better convergence accuracy in the new environment. The better performance of QLHPSO also shows its stronger robustness to the various dynamic scenarios in different instances.

Particularly, QLHPSO performs better than the comparison algorithms in the real-world application (instance JT56-DT3). That is to say, it can reschedule a near global optimal vehicle routing solution with lower transportation cost (including the fixed cost and the fuel cost) and less carbon emissions (the carbon emission cost) in reaction to the dynamic event. This demonstrates the practical utility of the proposed algorithm QLHPSO.

VI. CONCLUSION

There are two significant advantages of this work with the aim of dealing with the dynamically changing waste collection environment in real-world scenarios. The first one is that an energy-efficient multi-trip dynamic vehicle routing model is constructed for the waste collection problem. Two kinds of dynamic events that often occur during waste collection are introduced to address the dynamic characteristics of the environment. In order to increase the economic benefits while protecting environments, both the transportation cost and the carbon emissions are minimized, subjecting to the constraints of drivers' working hours and the vehicle capacity. Besides, each vehicle is allowed to have multiple trips to increase the resource utilization.

The second one is to propose a predictive-rescheduling method based on Q-learning-based hyperheuristic particle swarm optimization, termed QLHPSO, to solve the established model. Considering the problem features, including dynamically changing environments, two strongly coupled subproblems, complex constraints and multiple trips, four strategies are designed: 1) the encoding and decoding approach applicable for model is presented, which can simultaneously handle the two subproblems, while guaranteeing feasibility of the decoded solutions; 2) by adopting the dynamic response mechanism DRM, some high-quality individuals, including schedule repair solutions and historical solutions, are incorporated into the initial population to provide a guidance for the search of the algorithm at the early stage; 3) eight LLHs with different search functions are devised by combining four learning operators with two enhanced local search operators, which enriches the behavior mode of particles; and 4) four kinds of population states are defined based on convergence and diversity, and a Q-learning based high-level strategy QHLS is trained to learn to autonomously determine the most suitable LLH for the population with different states.

Besides, a comprehensive experimental study of the proposed algorithm QLHPSO is carried out. Two groups of experiments are performed on one real-world case and nine synthetic instances with different sizes. The first group performs an ablation study, where effectiveness of DRM, definition of the population state and QHLS is respectively validated. Experimental results indicate that the solution accuracy obtained by each new strategy is significantly better than that of the comparison ones on most instances, suggesting the feasibility and effectiveness of them. The second group compares QLHPSO with four state-of-the-art meta-heuristic algorithms to assess its overall performance. Experimental results demonstrate that QLHPSO outperforms all comparison algorithms with statistical significance in terms of the convergence accuracy. Once a dynamic event

occurs, it can react to it quickly and reschedule a near global optimal vehicle routing solution with lower transportation cost and less carbon emissions in the new environment. Moreover, QLHPSO has a good scalability to the problem size.

Although two kinds of dynamic events, working hours of the drivers and multiple trips of each vehicle are introduced in our model, some limitations still exist in our present study. First, it is still far from capturing all the factors, uncertainties and dynamic events which might affect the real waste collection. For instance, our current work adopts an average vehicle speed and assumes that the drivers can continue to work without respite. Nevertheless, the vehicle speed varies with road conditions, and the drivers need a break time at noon. In the future, more actual factors should be modeled. Second, more efficient LLHs could be devised, which provide a diverse set of candidates for the QHLS to choose, and further improve the search ability of our algorithm.

REFERENCES

[1] F. R. McDougall, P. R. White, M. Franke, and P. Hindle, *Integrated Solid Waste Management: A Life Cycle Inventory*. Hoboken, NJ, USA: Wiley, 2008.

[2] E. C. Rada, M. Grigoriu, M. Ragazzi, and P. Fedrizzi, "Web oriented technologies and equipments for MSW collection," in *Proc. Int. Conf. Risk Manag., Assessment Mitigation*, 2010, vol. 10, no. 201, pp. 150–153.

[3] T. E. Kanchanabhan, J. A. Mohaideen, S. Srinivasan, and V. L. K. Sundaram, "Optimum municipal solid waste collection using geographical information system (GIS) and vehicle tracking for Pallavapuram municipality," *Waste Manag. Res.*, vol. 29, no. 3, pp. 323–339, 2011.

[4] B. G. Wilson and B. W. Baetz, "Modeling municipal solid waste collection systems using derived probability distributions. I: Model development," *J. Environ. Eng.*, vol. 127, no. 11, pp. 1031–1038, Nov. 2001.

[5] V. Wagner and D. Rutherford, "Survey of best practices in emission control of in-use heavy-duty diesel vehicles," *Int. Council Clean Transp.*, vol. 64, pp. 64–116, Jul. 2013.

[6] S. Mousavi, A. Hosseinzadeh, and A. Golzary, "Challenges, recent development, and opportunities of smart waste collection: A review," *Sci. Total Environ.*, vol. 886, May 2023, Art. no. 163925.

[7] J. Luo and M.-R. Chen, "Multi-phase modified shuffled frog leaping algorithm with extremal optimization for the MDVRP and the MDVRPTW," *Comput. Ind. Eng.*, vol. 72, pp. 84–97, Jun. 2014.

[8] G. Laporte, "The vehicle routing problem: An overview of exact and approximate algorithms," *Eur. J. Oper. Res.*, vol. 59, no. 3, pp. 345–358, Jun. 1992.

[9] Z. Pan, L. Wang, C. Dong, and J.-F. Chen, "A knowledge-guided end-to-end optimization framework based on reinforcement learning for flow shop scheduling," *IEEE Trans. Ind. Informat.*, vol. 20, no. 2, pp. 1853–1861, Jun. 2023.

[10] S. Wang, Y. Wang, Y. Wang, and Z. Wang, "Comparison of multi-objective evolutionary algorithms applied to watershed management problem," *J. Environ. Manag.*, vol. 324, Dec. 2022, Art. no. 116255.

[11] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. Int. Conf. Neural Netw.*, 1995, pp. 1942–1948.

[12] X. Shen, H. Pan, Z. Ge, W. Chen, L. Song, and S. Wang, "Energy-efficient multi-trip routing for municipal solid waste collection by contribution-based adaptive particle swarm optimization," *Complex Syst. Model. Simul.*, vol. 3, no. 3, pp. 202–219, Sep. 2023.

[13] R. J. Kuo, M. Fernanda Luthfiansyah, N. Aini Masruroh, and F. Eva Zulvia, "Application of improved multi-objective particle swarm optimization algorithm to solve disruption for the two-stage vehicle routing problem with time windows," *Exp. Syst. Appl.*, vol. 225, Sep. 2023, Art. no. 120009.

[14] M. A. Islam, Y. Gajpal, and T. Y. ElMekkawy, "Hybrid particle swarm optimization algorithm for solving the clustered vehicle routing problem," *Appl. Soft Comput.*, vol. 110, Oct. 2021, Art. no. 107655.

[15] P. Cowling, G. Kendall, and E. Soubeiga, "A hyperheuristic approach to scheduling a sales summit," in *Proc. Int. Conf. Pract. Theory Automated Timetabling*, 2001, pp. 176–190.

[16] M. Sánchez, J. M. Cruz-Duarte, J. C. Ortíz-Bayliss, H. Ceballos, H. Terashima-Marin, and I. Amaya, "A systematic review of hyper-heuristics on combinatorial optimization problems," *IEEE Access*, vol. 8, pp. 128068–128095, 2020.

[17] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Mach. Learn.*, vol. 8, pp. 279–292, May 1992.

[18] R. S. Sutton and A. G. Barto, "Reinforcement learning: An introduction," *IEEE Trans. Neural Netw.*, vol. 9, no. 5, p. 1054, Sep. 1998.

[19] X.-N. Shen, L. L. Minku, N. Marturi, Y.-N. Guo, and Y. Han, "A Q-learning-based memetic algorithm for multi-objective dynamic software project scheduling," *Inf. Sci.*, vol. 428, pp. 1–29, Feb. 2018.

[20] Y. Yao, Z. Peng, and B. Xiao, "Parallel hyper-heuristic algorithm for multi-objective route planning in a smart city," *IEEE Trans. Veh. Technol.*, vol. 67, no. 11, pp. 10307–10318, Nov. 2018.

[21] D. Falcão, A. Madureira, and I. Pereira, "Q-learning based hyper-heuristic for scheduling system self-parameterization," in *Proc. 10th Iberian Conf. Inf. Syst. Technol. (CISTI)*, Jun. 2015, pp. 1–7.

[22] E. J. Beltrami and L. D. Bodin, "Networks and vehicle routing for municipal waste collection," *Networks*, vol. 4, no. 1, pp. 65–94, Jan. 1974.

[23] A. Louati, L. H. Son, and H. Chabchoub, "Smart routing for municipal solid waste collection: A heuristic approach," *J. Ambient Intell. Humanized Comput.*, vol. 10, no. 5, pp. 1865–1884, May 2019.

[24] N. A. Mat, A. M. Benjamin, and S. Abdul-Rahman, "Enhanced heuristic algorithms with a vehicle travel speed model for time-dependent vehicle routing: A waste collection problem," *J. Inf. Commun. Technol.*, vol. 17, pp. 55–78, Jan. 2017.

[25] E. Tirkolaee, A. Hosseinabadi, M. Soltani, A. Sangaiah, and J. Wang, "A hybrid genetic algorithm for multi-trip green capacitated arc routing problem in the scope of urban services," *Sustainability*, vol. 10, no. 5, p. 1366, Apr. 2018.

[26] Y. Shi, L. Lv, F. Hu, and Q. Han, "A heuristic solution method for multi-depot vehicle routing-based waste collection problems," *Appl. Sci.*, vol. 10, no. 7, p. 2403, Apr. 2020.

[27] H. Zafor, N. Mazumdar, and A. Nag, "An energy efficient waste collection, segregation, and processing of municipal solid waste in urban area using WSN," in *Proc. IEEE 9th Uttar Pradesh Sect. Int. Conf. Electr., Electron. Comput. Eng. (UPCON)*, Dec. 2022, pp. 1–6.

[28] Y. Shen, L. Yu, and J. Li, "Robust electric vehicle routing problem with time windows under demand uncertainty and weight-related energy consumption," *Complex Syst. Model. Simul.*, vol. 2, no. 1, pp. 18–34, Mar. 2022.

[29] M. L. Tee and D. E. Cruz, "A vehicle routing problem in plastic waste management considering the collection point location decisions," in *Proc. IEEE Int. Conf. Ind. Eng. Eng. Manag. (IEEM)*, Dec. 2022, pp. 551–555.

[30] J. C. Molina, I. Eguia, and J. Racero, "Reducing pollutant emissions in a waste collection vehicle routing problem using a variable neighborhood Tabu search algorithm: A case study," *TOP*, vol. 27, no. 2, pp. 253–287, Jul. 2019.

[31] H. Wu, F. Tao, and B. Yang, "Optimization of vehicle routing for waste collection and transportation," *Int. J. Environ. Res. Public Health*, vol. 17, no. 14, p. 4963, Jul. 2020.

[32] M. Erdem, "Optimisation of sustainable urban recycling waste collection and routing with heterogeneous electric vehicles," *Sustain. Cities Soc.*, vol. 80, May 2022, Art. no. 103785.

[33] H. Pamukçu, P. S. Yapicioglu, and M. I. Yesilnacar, "Investigating the mitigation of greenhouse gas emissions from municipal solid waste management using ant colony algorithm, Monte Carlo simulation and LCA approach in terms of EU green deal," *Waste Manag. Bull.*, vol. 1, no. 2, pp. 6–14, Sep. 2023.

[34] S. Hu, L. An, and L. Shen, "A multi-objective modeling and optimization approach to municipal solid waste collection for classified treatment in China towards sustainable development," *Sustain. Cities Soc.*, vol. 98, Nov. 2023, Art. no. 104846.

[35] H. R. Lewis, R. G. Michael, and D. S. Johnson, "Computers and intractability. A guide to the theory of NP-completeness," *J. Symbolic Log.*, vol. 48, no. 2, pp. 498–500, 1983.

[36] M. J. Cortinhal, M. C. Mourão, and A. C. Nunes, "Local search heuristics for sectoring routing in a household waste collection context," *Eur. J. Oper. Res.*, vol. 255, no. 1, pp. 68–79, Nov. 2016.

[37] M. Akhtar, M. A. Hannan, R. A. Begum, H. Basri, and E. Scavino, "Backtracking search algorithm in CVRP models for efficient solid waste collection and route optimization," *Waste Manag.*, vol. 61, pp. 117–128, Mar. 2017.
[38] X. Jin, H. Qin, Z. Zhang, M. Zhou, and J. Wang, "Planning of garbage collection service: An arc-routing problem with time-dependent penalty cost," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 5, pp. 2692–2705, May 2021.
[39] G. Marseglia, J. A. Mesa, F. A. Ortega, and R. Piedra-De-La-Cuadra, "A heuristic for the deployment of collecting routes for urban recycle stations (eco-points)," *Socio-Economic Planning Sci.*, vol. 82, Aug. 2022, Art. no. 101222.
[40] S. Voß, S. Martello, I H. Osman, and C. Roucairol, *Meta-Heuristics: Advances and Trends in Local Search Paradigms for Optimization*. Dordrecht, The Netherlands: Kluwer, 2012.
[41] E. B. Tirkolaee, M. Alinaghian, M. B. Sasi, and M. M. S. Esfahani, "Solving a robust capacitated arc routing problem using a hybrid simulated annealing algorithm: A waste collection application," *J. Ind. Eng. Manag. Stud.*, vol. 3, no. 1, pp. 61–76, 2016.
[42] M. A. Hannan, M. Akhtar, R. A. Begum, H. Basri, A. Hussain, and E. Scavino, "Capacitated vehicle-routing problem model for scheduled solid waste collection and route optimization using PSO algorithm," *Waste Manag.*, vol. 71, pp. 31–41, Jan. 2018.
[43] Q. Wei, Z. Guo, H. C. Lau, and Z. He, "An artificial bee colony-based hybrid approach for waste collection problem with midway disposal pattern," *Appl. Soft Comput.*, vol. 76, pp. 629–637, Mar. 2019.
[44] N. Wichapa, T. Sudsuansee, and P. Khokhajaikiat, "Solving the vehicle routing problems with time windows using hybrid genetic algorithm with push forward insertion heuristic and local search procedure," *J. King Mongkut's Univ. Technol. North Bangkok*, vol. 29, no. 1, pp. 4–13, Dec. 2018.
[45] Q. Qiao, F. Tao, H. Wu, X. Yu, and M. Zhang, "Optimization of a capacitated vehicle routing problem for sustainable municipal solid waste collection management using the PSO-TS algorithm," *Int. J. Environ. Res. Public Health*, vol. 17, no. 6, p. 2163, Mar. 2020.
[46] K. Tomitagawa, A. Anuntachai, S. Chotiphan, O. Wongwirat, and S. Kuchii, "Adapted ACO algorithm for energy-efficient path finding of waste collection robot," in *Proc. 22nd Int. Conf. Control, Autom. Syst. (ICCAS)*, Nov. 2022, pp. 469–474.
[47] T. Bektas and G. Laporte, "The pollution-routing problem," *Transp. Res. B, Methodol.*, vol. 45, no. 8, pp. 1232–1250, Sep. 2011.
[48] I. Gölcük and F. B. Ozsoydan, "Q-learning and hyper-heuristic based algorithm recommendation for changing environments," *Eng. Appl. Artif. Intell.*, vol. 102, Jun. 2021, Art. no. 104284.
[49] I. Sbai, S. Krichen, and O. Limam, "Two meta-heuristics for solving the capacitated vehicle routing problem: The case of the Tunisian post office," *Oper. Res.*, vol. 22, no. 1, pp. 507–549, Mar. 2022.
[50] J. J. Liang, A. K. Qin, P. N. Suganthan, and S. Baskar, "Comprehensive learning particle swarm optimizer for global optimization of multimodal functions," *IEEE Trans. Evol. Comput.*, vol. 10, no. 3, pp. 281–295, Jun. 2006.
[51] S. Hougardy, F. Zaiser, and X. Zhong, "The approximation ratio of the 2-Opt heuristic for the metric traveling salesman problem," *Oper. Res. Lett.*, vol. 48, no. 4, pp. 401–404, 2020.
[52] S. Gülcü, M. Mahi, Ö. K. Baykan, and H. Kodaz, "A parallel cooperative hybrid method based on ant colony optimization and 3-Opt algorithm for solving traveling salesman problem," *Soft Comput.*, vol. 22, no. 5, pp. 1669–1685, Mar. 2018.
[53] The VRP Web. (2016). *CVRP Instances*. Accessed: Nov. 6, 2022. [Online]. Available: http://www.bernabe.dorronsoro.es/vrp/
[54] J. Chen, B. Dan, and J. Shi, "A variable neighborhood search approach for the multi-compartment vehicle routing problem with time windows considering carbon emission," *J. Cleaner Prod.*, vol. 277, Dec. 2020, Art. no. 123932.
[55] M. Moodi, M. Ghazvini, and H. Moodi, "A hybrid intelligent approach to detect Android botnet using smart self-adaptive learning-based PSO-SVM," *Knowl.-Based Syst.*, vol. 222, Jun. 2021, Art. no. 106988.
[56] N. R. Sabar, A. Bhaskar, E. Chung, A. Turky, and A. Song, "A self-adaptive evolutionary algorithm for dynamic vehicle routing problems with traffic congestion," *Swarm Evol. Comput.*, vol. 44, pp. 1018–1027, Feb. 2019.
[57] G. Xue, Y. Wang, X. Guan, and Z. Wang, "A combined GA-TS algorithm for two-echelon dynamic vehicle routing with proactive satellite stations," *Comput. Ind. Eng.*, vol. 164, Feb. 2022, Art. no. 107899.
[58] M. Okulewicz and J. Mandziuk, "A metaheuristic approach to solve dynamic vehicle routing problem in continuous search space," *Swarm Evol. Comput.*, vol. 48, pp. 44–61, Aug. 2019.
[59] M. Okulewicz and J. Mandziuk, "The impact of particular components of the PSO-based algorithm solving the dynamic vehicle routing problem," *Appl. Soft Comput.*, vol. 58, pp. 586–604, Sep. 2017.

YUN ZHAO received the B.S. degree from Anhui Agriculture University and the Ph.D. degree from Nanjing University of Science and Technology. She is currently a Lecturer with the School of Electrical Engineering, Nanjing Vocational University of Industry Technology, Nanjing, China. Her research interests include computational intelligence, reinforcement learning, and multi-objective optimization.

XIAONING SHEN received the B.S. and Ph.D. degrees from Nanjing University of Science and Technology, in 2003 and 2008, respectively.

She is currently a Professor with the School of Automation, Nanjing University of Information Science and Technology, Nanjing, China. She has published over 70 academic articles in the area of evolutionary algorithms and their applications. Her research interests include computational intelligence, and multi-objective optimization and scheduling.

WENYAN CHEN received the B.S. degree from the Xuhai College, China University of Mining and Technology. She is currently pursuing the master's degree with Nanjing University of Information Science and Technology, Nanjing, China. Her research interest includes hyperheuristic algorithm and its application in software project scheduling.

HONGLI PAN received the B.S. degree from Wanjiang University of Technology and the master's degree from Nanjing University of Information Science and Technology, Nanjing, China. Her research interests include swarm intelligence algorithms and their applications.