Research article

Multi-task dispatch of shared autonomous electric vehicles for Mobility-on-Demand services – combination of deep reinforcement learning and combinatorial optimization method

Ning Wang, Jiahui Guo*

School of Automotive Studies, Tongji University, Shanghai, 201804, China

ARTICLE INFO

Keywords:
 Sustainable transportation
 Recharging task assignment
 Repositioning task assignment
 Bipartite graph matching
 Maximum flow problem

ABSTRACT

The Autonomous Mobility-on-Demand system is an emerging green and sustainable transportation system providing on-demand mobility services for urban residents. To achieve the best recharging, delivering, and repositioning task assignment decision-making process for shared autonomous electric vehicles, this paper formulates the fleet dynamic operating process into a multi-agent multi-task dynamic dispatching problem based on Markov Decision Process. Specifically, the decision-making process at each time step is divided into 3 subprocesses, among which recharging and delivery task assignment processes are transformed into a maximum weight matching problem of bipartite graph respectively, and the repositioning task assignment process is quantified as a maximum flow problem. Kuhn-Munkres Algorithm and Edmond-Karp Algorithm are adopted to solve the above two mathematical problems to achieve the optimal task allocation policy. To further improve the dispatching performance, a new instant reward function balancing order income with trip satisfaction is designed, and a state-value function estimated by Back Propagation-Deep Neural Network is defined as a matching degree between each shared autonomous electric vehicle and each delivery task. The numerical results show that: (i) a reward function focusing on income and satisfaction can increase total revenue by 33.2%, (ii) the introduction of task allocation repositioning increases total revenue by 50.0%, (iii) a re-estimated state value function leads to a 2.8% increase in total revenue, (iv) the combination of charging and task repositioning can reduce user waiting time and significantly improve user satisfaction with the trip.

1. Introduction

Global urban transportation is facing booming revolutions including sharing, electrification, and automation. A Mobility-on-Demand system is being formed by the combination of autonomous driving, electric vehicles, and car-sharing mode. In the Autonomous Mobility-on-Demand (MoD) system, Shared Autonomous Electric Vehicles (SAEVs) are the main fleet that executes tasks including passenger delivery from origin to destination, recharging when the electricity is not enough, repositioning from oversupply sub-regions to undersupply sub-regions, and parking. On the one side, SAEVs assist cities cut carbon emissions, pollution, and traffic congestion. On the other side, SAEVs may provide daily mobility services to satisfy the demands of consumers on their trips.

The paper provides new insights into the dynamic multi-agent multi-task relocation problem for Shared Autonomous Electric Vehicles. A particular focus is on the best decision procedure for allocating SAEVs to meet the daily trip order of users, to target charging stations to replenish energy, and to travel to underserved sub-regions to replenish vehicles ahead of time. China is chosen to be the target research area because electric and intelligent vehicles develop the fastest in the world. As of 2021, Baidu Apollo has completed more than 16 million kilometers of Robotaxi fleet driving tests in China. Because it is a relatively new issue that has emerged in recent years, few studies have previously concentrated on the relocation challenge of SAEVs.

The majority of the issues studied so far focus on the dispatching problem of personal-based electric car sharing systems (Aitouahmed et al., 2017; Shaheen and Cohen, 2013; Wang et al., 2020). However, these researches almost entirely concentrate on the user's delivery task, ignoring the problem of charging electric vehicles and the rebalancing problem of superfluous vehicles (Nourinejad and Roorda, 2014; Nourinejad et al., 2015, Xu et al., 2018a,b). Meanwhile, in the experi-

* Corresponding author.
E-mail addresses: wangning@tongji.edu.cn (N. Wang), gjh950719@tongji.edu.cn (J. Guo).

https://doi.org/10.1016/j.heliyon.2022.e11319
Received 19 April 2022; Received in revised form 6 June 2022; Accepted 24 October 2022

2405-8440/© 2022 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
ments mentioned above, nonlinear programming mathematical models and solution methods with long calculation times are applied, which neglect computational efficiency in humongous applications (Boyac et al., 2015; De Almeida Correia and Antunes, 2012; Hu et al., 2019). As a result, present approaches are incompatible with the future MoD system. There is a pressing need for a novel system for humongous multi-task relocation of SAEVs that takes delivering, recharging, and repositioning tasks into account.

Instead of static dispatching, this paper intends to simulate the decision-making procedure of the MoD system, and optimizes the multi-task dynamic relocation of SAEVs throughout the course of a long period of time (e.g. one hour). A mathematical modeling method is used to depict the entire multi-task relocation procedure of SAEVs. To provide the optimum task allocation result for SAEVs in large-scale applications, an optimization model is built to establish the goals and constraints, as well as a rapid solution technique.

Therefore, at first, the Markov Decision Process (MDP) framework (Mao et al., 2020) is adopted to model the multi-agent multi-task dynamic dispatching problem of SAEVs including agent, state, action, and reward. Then the multi-task dispatching process at each time step is divided into 3 sub-processes, among which recharging and delivery task assignment processes are transformed into a maximum weight matching problem of bipartite graph respectively, and the repositioning task assignment process is quantified as a maximum flow problem. Kuhn-Munkres Algorithm (Zhu et al., 2016) and Edmond-Karp Algorithm (Lammich and Sefidgar, 2016) are adopted to solve the above two mathematical problems to provide the most effective task allocation strategies. After that, based on the Bellman Equation (Jones and Peet, 2021), a new state-value function is developed to change the single-step delivery reward into a multi-step return which represents as the matching degree between the delivery task and the SAEV. The Back Propagation-Deep Neural Network (BP-DNN) algorithm (Zheng et al., 2022) is adopted to estimate the state-value function based on historical trip episode samples. Finally, 4 simulation test cases are designed to verify the operational performance of the above methodology. Numerical results show some impressive findings compared with previous studies, and it is proved that the combination of combinatorial optimization method and deep reinforcement learning can effectively solve multi-task dynamic dispatching problem of SAEVs in the future.

The proposed methods in this paper are appropriate for the large-scale deployment of a future fleet of SAEVs both in the aspect of operational performance and computational efficiency. The contributions of this study, also the research objectives of this paper, are summarized below:

1. Combining the delivery reward function that balances user waiting time and order revenue with a BP-DNN based state-value function that concerns long-term return can bring a better operating performance including user waiting time reduction and order revenue improvement.
2. Introducing repositioning task assignment and BP-DNN-based state-value function into the dynamic dispatching process can help SAEVs fleet operators increase total order revenue and decrease the average user waiting time to a great extent.
3. Breaking the multi-task dynamic dispatching process at each timestep into 2 bipartite graph maximum weight matching problems (\(\alpha n^2\)) and 1 maximum flow problem (\(OV^*E^2\)) can reduce the complexity of the solution to a large extent in practical application.

The theoretical contributions of this paper mainly include two aspects. First, to optimize the decision-making process of multi-task dynamic relocation of SAEVs, a theoretical optimization model is constructed based on deep reinforcement learning, optimal matching of bipartite graphs, and network maximum flow methods. Second, to solve the optimization model, a solving algorithm with low-space-time complexity based on the Kuhn-Munkres algorithm and the Edmond-Karp algorithm is proposed.

The remainder of the paper is laid out as follows. Section 2 provides a survey of the literature on the dispatching challenge in the Autonomous Mobility-on-Demand system. The analytical framework is presented in Section 3 along with the methodology, which includes problem description, problem assumption, and mathematical formulation. Section 4 lays out the policies that will be used to solve the aforementioned mathematical difficulties. The numerical results of four simulation test scenarios are shown in Section 5 to validate the proposed methodology’s performance. Section 6 concludes the research findings and looks ahead to future work.

2. Literature review

Trip demand, route planning, fleet size, vehicle assignment, pricing, charging, vehicle distribution, and parking are 8 common sub-problems in the fleet operating process of SAEVs. (Narayanan et al., 2020). This research is conducted based on assumptions of certain traffic assignment, trip demand, fleet size, and pricing. Trip demand and fleet size are achieved based on the existing fleet configuration and trip orders in the Didi platform (https://outreach.didichuxing.com). Pricing and route planning including route planning and navigation scheme follow the “real-time shortest driving routing” mechanism conducted by the AutoNavi Map platform (https://lbs.amap.com/api/webservice/guide/api/direction).

Instead, when studying the multi-agent multi-task dynamic relocation problem of SAEVs in the Mobility-on-Demand system, the main attention in this research is trip order assignment, vehicle parking and charging assignment, and vehicle rebalancing. Most scholars examine these three sub-problems independently (as indicated in Table 1), this paper combines combinatorial optimization methods and deep reinforcement learning to tackle all three sub-problems in one methodology framework.

Although our previous research Wang and Guo (2021) is the first to model these three sub-problems in one framework by the reinforcement learning method, this study differs in 3 aspects: (1) Although the fleet dispatching process is also modeled by the MDP framework, the reward function of delivery action in this paper is improved by considering both order revenue and user satisfaction; (2) Three sub-problems are transferred into a 3-stage decision process rather than a 1-stage decision process, in which recharging and delivery task assignment are transferred into two individual bipartite graph matching problems separately, and repositioning task assignment is transferred into a max flow problem; (3) When designing the solving algorithm, Edmond-Karp algorithm is first adopted except for Kuhn-Munkres algorithm and deep Q-learning algorithm.

2.1. Trip order assignment

Trip order assignment entails allocating some vehicles to users, after which these vehicles perform pickup and delivery tasks to fulfill the users’ trip requests. The three major strategies for solving this problem are rules, heuristics, and a precise optimization algorithm. The rule-based vehicle assignment technique is typically used to simulate the dynamic process of vehicle assignment (Chen and Kockelman, 2016; Fagnant and Kockelman, 2015, Gurumurthy and Kockelman, 2018; Hörl, 2017; Jäger et al., 2017, Mendes et al., 2017; Zhang and Guhathakurta, 2017). The most common rule is to assign the user’s request to the nearest vehicle. To improve vehicle assignment performance, more optimization mathematical models and solving methods are being created. To achieve the best assignment plan, the dial-a-ride problem of ride-sharing SAEVs in the urban road network is modeled as an integer non-linear programming problem by Liang et al. (2020). To solve the optimization problem, they present a modified Lagrangian relaxation approach. However, due to the calculation time and the computation gap between the lower and upper bounds, this methodology is not suitable for use in practice. This is a common issue in other studies as well (Braekers et al., 2014, Farhan and Chen, 2018; Martinez and Viegas, 2017; Paquette et al., 2013). Shi et al. (2019) present a methodology
for operating an electric vehicle fleet based on a reinforcement learning method, which may be used for the trip order assignment problem of SAEVs. The goal of creating a reward function is to minimize customer waiting time, economic impact, and electricity costs. To approximate the state-value function, a deep feed-forward neural network is used, and a KM algorithm with a time complexity of $o(n^3)$ is employed to discover the best dispatching outcomes.

### 2.2. Parking and recharging assignment

Monitoring the real-time state of charge (SOC) of SAEVs and implementing related techniques to appoint some vehicles to a specific charging station/pile or parking space is referred to as parking and recharging assignment (Iglesias et al., 2018). According to Chen and Kockelman (2016), while charging cars are not allowed to undock and service a new trip request, Bauer et al. (2018) think that still-charging vehicles are. In a mixed-integer optimization model, Iacobucci et al. (2019a,b) employ energy pricing information to optimize car charging by introducing some recharging restrictions across longer periods. Jones and Leibowitz (2019) use a mathematical model for energy optimization to assess the impact of charging SAEVs at periods when the energy system is most efficient. To locate parking spaces where vehicle recharging is also viable, Azevedo et al. (2016) apply an optimization approach (Facility Location issue). By moving idling vehicles to the parking areas at a low cost, Zhang and Guhathakurta (2017) can minimize operational costs. Al-Kanj et al. (2020) integrate SAEVs into the ride-hailing systems, and use a combination of an on-stage optimization method and the Markov decision process to achieve real-time optimal parking and recharging assignment decision-making.

### 2.3. Vehicle rebalancing

When modeling the on-demand mobility service based on SAEVs, vehicle rebalancing, also known as “the vehicle repositioning or redistribution”, is utilized to move vehicles from oversupply regions to undersupply regions. For SAEV operations, Fagnant and Kockelman (2014) develop an agent-based simulation model that includes vehicle rebalancing. According to the findings, vehicle rebalancing might save 10 times the number of vehicles required for personal transport. According to Vosooghi et al. (2019), vehicle rebalancing has a significant influence on mobility service performance, including modal share and fleet utilization. Babicheva et al. (2018), on the other hand, adopt six alternative methods for assigning vehicle rebalancing tasks and propose a new index-based proactive redistribution (IBR) algorithm based on predicted near-future demand at stations. Zhang and Pavone (2014) and Alonso-Mora et al. (2017) use a method of linear programming for vehicle repositioning. The challenges of trip order assignment and repositioning vehicles can be decoupled, according to Rossi et al. (2016), who build an SAEV routing and computationally efficient repositioning algorithm, and the vehicle repositioning optimization issue is described as a Minimum Flow problem. Dandl et al. (2019) stress the importance of trip order demand forecasting, claiming that high-quality predictions can help with vehicle redistribution. Mao et al. (2020) adopt a Markov Decision Process to simulate the vehicle rebalancing task relocation problem of SAEVs, and an actor-critic strategy gradient network to get the optimal dispatching results.

### 2.4. Research gaps

In general, existing research into the multi-agent multi-task dynamic dispatching problem of SAEV operation is insufficient, and three kinds of research gaps must be filled.

First, the past study has generally studied trip order assignment, recharging, parking assignment, and vehicle rebalancing independently. When modeling and solving these problems, a few scholars integrate vehicle assignment and charging assignment. However, because vehicle redistribution has a substantial impact on the performance of mobility services, it is preferable to incorporate vehicle redistribution task assignment into the decision-making procedure throughout the dynamic operating process.

Second, when choosing which vehicle should be allocated to a specific passenger, customer satisfaction, operational income, and power cost are barely balanced in the earlier studies. Furthermore, constructing each SAEV-passerenger pair’s matching weight from a long perspective has been proven to be superior to a short-sighted instant perspective. As a result, creating a multi-dimensional and cumulative reward to indicate the matching degree of each SAEV and delivery task is desirable.

Third, combining the combinatorial optimization method and deep reinforcement learning is a viable methodology framework for ensuring that these methods are practical-ready for SAEVs’ dynamic dispatching process. Some studies have shown that the aforementioned methodology framework is effective, although they are exclusively used in the ride-hailing industry rather than in the operation of SAEV fleets.
way to model the dynamic relocation process using a reinforcement learning framework that includes delivery, charging, and repositioning tasks, as well as the way to solve the best recharging, trip order assignment (delivery), and rebalancing assignment result using combinatorial optimization methods, are still being investigated.

3. Multi-task dynamic assigning problem of SAEVs

The MoD system is composed of an autonomous fleet of battery electric vehicles (BEVs) that provide daily mobility services for urban residents to meet their daily trip demand. One of the main tasks of the MoD system is to respond to each user’s trip request and then assign one vehicle to pick up and deliver the user to the destination. Also, since the fleet is powered by electricity, recharging is another key task to be assigned when the state of charge (SOC) of BEV is at a low battery level. Besides that, due to the variation of vehicle supply and trip order demand in different regions and time periods, supply-demand imbalance happens in high frequency, hence vehicle repositioning task for regional rebalance is necessary. In addition to the above three tasks, the parking task will be assigned to the remaining vehicles waiting for new tasks, and those remaining vehicles are assumed to stay in the original position in this paper.

Therefore, the operation process of the SAEVs fleet in the MoD system can be defined as the decision-making process of a multi-task dynamic assigning problem mainly involved with four tasks including delivery, recharging, repositioning, and parking. This work aims to design the optimal task allocation policy in real-time for SAEVs considering order revenue, user satisfaction, and energy cost. To realize this, a three-stage multi-task dynamic assigning method combining Kuhn-Munkres (KM) algorithm, Edmond-Karp (EK) algorithm, and Back Propagation Deep Neural Network is proposed in this paper.

3.1. Problem assumption

The multi-task dynamic assigning problem in this paper is modeled using Markov Decision Process (MDP). In the MDP framework, each SAEV is viewed as an agent, and the 1st assumption is that no manpower is involved in the whole decision process. Specifically, each SAEV executes tasks autonomously without a human driver, and the recharging task is assumed to be completed by the SAEV itself without human intervention.

The 2nd assumption of this study is that the forecast models of estimated travel time (ETT), estimated travel distance (ETD), and estimated energy consumption (EEC) for each pair of Origin-Destination requests are known respectively. In this model, when the origin, the waypoint, and the destination are known, the reward of choosing each action (recharging, delivery, and repositioning) considering order revenue, user satisfaction, and energy cost is designed according to estimated travel time (EPT) (i.e. user waiting time), estimated travel time, estimated travel distance, and estimated energy consumption. EPT, ETT, ETD, and EEC of each action are presumed to be calculated based on the traffic assignment policy of the AutoNavi Map platform, which follows the principle of the real-time shortest driving route. EEC is hypothesized to be achieved by constructing a new estimation model by combining long-short term memory network (LSTM) and random forest (RF) based on the real-time operational parameters of the vehicle itself such as velocity, acceleration, soc, temperature, etc. (Wang et al., 2019).

The 3rd assumption is that all the operational regions are pre-divided and predetermined in the form of hexagonal lattices, and the trip demand forecast model has been established to predict whether some region is oversupply or undersupply at each time period. To judge whether a region is oversupply or undersupply, the predicted trip demand will be compared with the current vehicle supply number. The related forecast work can refer to Wang et al. (2019). Therefore, the absent vehicles in undersupply and the residual vehicles in oversupply regions at each time period are known in this research.

3.2. Problem formulation

The MDP framework of the multi-task dynamic assigning problem of SAEVs consists of 7 core components: state, action, reward function, transition function, state-value function, decision variables, and an objective function. Each SAEV is modeled as an agent. The flow chart of the dispatching problem modeling process is shown in Fig. 1.

State

The state of each agent consists of temporal, spatial, and energy information. The temporal state, $t$, represents the Unix timestamp of each agent. The spatial state, $(lng, lat)$, describes the real-time geographical location of each agent. $lng$ and $lat$ represent the longitude and latitude of the agent respectively. The energy state, $e$, denotes the real-time remaining electricity of the battery, which is calculated by subtracting the amount of electricity consumed between the previous and current state from the amount of remaining electricity in the previous state. The electricity consumed between two states $E_{EC}$ is calculated by multiplying the average power consumption of 100 kilometers (power loss/100 km) with travel distance $ETD$. In other words, the state of each agent is implied by a vector $(t, lng, lat, e)$.

Action

Each agent executes four actions, including recharging $a_1$, delivery $a_2$, repositioning $a_3$, and parking $a_4$. Recharging denotes the SAEV should head for the assigned charging pile for power replenishment. Delivery means the SAEV drives to the origin to pick up and drop off the passenger at their desired location. Repositioning implies the SAEV drives to the undersupply region from the oversupply region for vehicle replenishment. Parking denotes the SAEV stays in the original position if no other tasks are assigned.

Reward function

For each agent, choosing any action will bring an expected reward to the whole MoD system. The principle of designing the reward should consider order revenue, user satisfaction, and energy cost simultaneously. Specifically, order revenue is directly related to $ETT$ and $ETD$ of each pair of Origin-Destination (O-D). User satisfaction is relevant to estimated pickup time ($EPT$ or user waiting time), which represents the estimated elapsed time of the pickup process from where the agent locates to where the passenger locates. The longer the pickup process is, the lower the user satisfaction is. Energy cost indicates the expenditure of the charging fee, which is linearly related to $E_{EC}$.

In the algebraic form, for the O-D pair, the reward $r_1$ of action recharging $a_1$ can be written as Equation (1). Vehicles are encouraged to be recharged nearby. Hence, $r_1$ is set as a negative of the recharging task’s $E_{EC}$, which is linearly related to $ETD$. The higher $ETD$ is, the less reward $r_1$ is. $x$ represents a linear factor. The reward $r_2$ of action delivery $a_2$ can be written as Equation (2). Based on the estimated order income (linearly related to $ETT$ and $ETD$, $a$ and $\beta$ are linear correlation coefficients related to $ETT$ and $ETD$ respectively), a satisfaction discount factor $\Omega$ is set to expand or reduce the reward considering the user trip satisfaction level represented by user waiting time $EPT$. $\Omega$ is determined by Equation (3). Set acceptable waiting time ($AWT$) as the boundary satisfaction level for users. If $EPT$ is greater than $AWT$, then $\Omega > 1$, which will reduce the reward of this action. If $EPT$ is less than $AWT$, then $\Omega < 1$, which will expand the reward of this action. Since repositioning action $a_3$ and parking action $a_4$ will not bring direct income for the MoD system, the reward of $a_3$ and $a_4$ is set as 0, as is shown in Equation (4).

$$r_1 = -E_{EC} = -x \times ETD$$  

(1)
Transition function

For an agent with the state \((i, l, e)\), any of these 4 actions will lead the agent to another state \((i', l', e')\). During each round of assigning process, for action \(a_1\), \(l'\) denotes the location of the target charging pile. For action \(a_2\), \(l'\) means the starting point of the trip request put forward by the user. For action \(a_3\), \(l'\) implies the center position of each hexagonal region with insufficient SAEVs. It is worth noting that once the agent enters the undersupply region, that is, crosses the boundary of the hexagonal region, this repositioning task can be finished and the state of this agent can be updated to the one waiting for the newest round of assignment. For action \(a_4\), \(l' = l\). To finish each task, an agent will take \(\Delta t\) minutes and consume \(\Delta e\) kWh, the formula is shown in Equation (5) and (6). It is noteworthy that all the agents cannot be assigned to new tasks only if they have finished the current task.

\[
\Delta t = l' - t = ETT \\
\Delta e = e' - e = EEC
\]

State-value function

Here a new state-value function is defined to represent the long-term utility of each state when making delivery task assignment decisions. On the one hand, if the matching degree is considered only in the local view, the instant reward of each pair of agents and task should be calculated by the above reward function following Equation (1)-(4). On the other hand, if the long-term return is considered, the matching degree ought to be quantified by adding the instant reward with a future utility by following Bellman Equation. Therefore, Equation (7) reveals the algebraic form of accumulated return for the agent executing the delivery task from a global view. \(V(s)\) represents the accumulated return for the agent at state \(s\), \(V(s')\) represents the accumulated return for the agent at state \(s'\), \(r_{2,2 \rightarrow s'}\) represents the instant return (following the equation of \(r_2\)) for the agent changing from state \(s\) to state \(s'\). The discount factor, \(\gamma \in (0, 1)\) (Xu et al., 2018a, b), determines how far the MDP framework can project into the future. Because extended horizons lead the state-value function to have a lot of volatility, it’s best to employ a small discount factor.

\[
V(s) = r_{2,2 \rightarrow s'} + \gamma \times V(s')
\]

The derivation process of the Bellman equation:

\[
V(s) = G_i \mid S_t = s \\
= R_{i,t} + \gamma R_{i,t+1} + \gamma^2 R_{i,t+2} + \cdots \mid S_t = s \\
= R_{i,t} + \gamma (R_{i,t+1} + \gamma R_{i,t+2} + \cdots) \mid S_t = s \\
= R_{i,t} + \gamma G_{i,t+1} \mid S_t = s \\
= R_{i,t} + \gamma V(S_{i,t+1}) \mid S_t = s
\]

Because:

\[
R_{i,t} = r_{2,2 \rightarrow s'} \quad V(S_{i,t+1}) = V(s')
\]

Hence:

\[
V(s) = r_{2,2 \rightarrow s'} + \gamma \times V(s')
\]

Explanation of discount factor \(\gamma\):

1. When \(\gamma = 0\), \(V(s) = r_{2,2 \rightarrow s'}\); (2) when \(\gamma = 1\), \(V(s) = r_{2,2 \rightarrow s'} + V(s') = r_{2,2 \rightarrow s'} + R_{i,t+1} + R_{i,t+2} + \cdots \); (3) when \(\gamma \in (0, 1)\), infinite reward \(R_{i,t+1}, R_{i,t+2}, R_{i,t+3}, \ldots\) will not be directly added to become the total reward, because the multiplication of infinite \(\gamma\) will tend to be 0. This treatment considers the discounting of future rewards.

Decision variables

Suppose the studied area is divided into \(M\) sub-regions. The radius of each sub-region is \(s km\). The whole dispatching process at each time period \(t\) mainly includes 2 parts. One is charging and delivery task assignment of each sub-region, the other is vehicle repositioning among all the sub-regions.

For each sub-region \(m\), the charging and delivery task assignment process at each time period \(t\) is as follows: First, all the available vehicles in sub-region \(m\) are assembled into a set \(I_{tm}^m\). Second, the number of vehicles with the electricity of less than \(\Lambda\) kWh is screened as \(I_{tm}^m\) and these vehicles will be assigned to \(J^m\) charging piles in this sub-region.

\[
r_2 = \frac{a \times ETT + b \times ETD}{\Omega}
\]

\[
\Omega = \frac{EPT}{AWT}
\]

\[
r_3 = r_1 = 0
\]
m. Hence, the first decision variable is \(x_{im}^m\), which means the decision for vehicle \(i_m^m\) to drive to charging pile \(j_m^m\) in sub-region \(m\). If the trip orders are assembled into a set \(O_m\) and these orders should be assigned to the remaining \((I_m^m - I_m^m)\) vehicles with the electricity of more than \(\Lambda\) kWh. All the available vehicles for executing delivery tasks are assembled into a set \(I_m^m\). Thus, the second decision variable is \(i_{jm}^m\), which means the decision for vehicle \(i_{jm}^m\) to execute a delivery task and complete the order \(o_j^m\) in \(O_m\).

After finishing the above 1st part of the dispatching process at time period \(t\), the number of remaining vehicles in each sub-region \(m\) can be counted as \(I_m^{t+1}\). The 2nd part of the dispatching process among all the sub-regions follows. First, the trip order demand volume of each sub-region \(m\) in the next time period \((t + 1)\) can be predicted as \(O_m^{t+1}\). By comparing \(O_m^{t+1}\) with \(I_m^{t+1}\), each sub-region \(m\) can be judged as oversupply \((I_m^{t+1} > O_m^{t+1})\), balance \((I_m^{t+1} = O_m^{t+1})\), or undersupply \((I_m^{t+1} < O_m^{t+1})\), and the status of each sub-region is recorded by variable \(S_m\). Second, all the oversupply sub-regions are screened as a set \(A\), and all the undersupply sub-regions are screened as a set \(B\). There is a shortage of \(D_m^m\) vehicles in undersupply sub-region \(m\) for the next round of the dispatching process \((D_m^m = O_m^{t+1} - I_m^{t+1})\). Each vehicle \(i_m^m\) in oversupply sub-regions will be decided whether to drive to the undersupply sub-region \(b\) in \(B\). Therefore, the third decision variable is \(x_{ib}^m\), which means the decision for vehicle \(i_m^m\) to drive to the sub-region \(b\).

It is noted that the available vehicles are defined as vehicles for delivery tasks. Since the status of each vehicle will be updated at each time period, the vehicle assigned to the repositioning task at time period \(t\) may be updated as available vehicles at time period \(t + 1\) because they may have finished the repositioning task and reached the undersupply region waiting for the delivery tasks. Hence, the repositioning vehicle can be considered as the available vehicle at a suitable time period.

**Objective function and constraints**

When assigning vehicles with low electricity to charging piles nearby at each time period, the optimization goal is to maximize the total recharging reward. \(f_{ij}^m\) represents the recharging reward if vehicle \(i_m^m\) drives to charging pile \(j_m^m\) for electricity replenishment. The objective function and constraints are shown as Equation (8)-(9).

\[
\text{Max} \ W = r_{ij}^m \cdot x_{ij}^m \\
\text{s.t.} \quad \sum_{i \in I_{2m}} x_{ij}^m = 1 \\
\sum_{j \in J_{2m}} x_{ij}^m = 1
\]

(8)

While assigning the other vehicles to different delivery tasks, the optimization goal is set to the total reward. \(r_{im}^m\) denotes the reward if vehicle \(i_m^m\) finishes order \(o_j^m\) (including pickup and delivery) and it is calculated by adding instant reward function \(r_2\) with state-value function \(V(s)\). The objective function and constraints are shown as Equation (10)-(11).

\[
\text{Max} \ R = r_{im}^m \cdot x_{im}^m \\
\text{s.t.} \quad \sum_{i \in I_{2m}} x_{im}^m = 1 \\
\sum_{o \in O_{2m}} x_{im}^m = 1
\]

(10)

The vehicle repositioning process can be modeled as a maximum flow problem. The directed connected graph is shown in Fig. 2. Accordingly, volume \(V\) is set as the optimization objective, \(f^*\) means one feasible flow from node \(s\) to node \(t\), \(f_{ij}\) denotes the flow from node \(i\) to node \(j\). \(V(f)\) illustrates the net output of node \(s\) or the net input of node \(t\), and it is equal to the feasible flow \(f^*\). The capacity of each edge from node \(s\) to node \(m\) which in “undersupply sub-regions” is marked as \(I_{im}^m\). Each node \(m\) can only be connected by node \(i\) in “available vehicles” which are in the corresponding sub-region and the capacity is set as 1. Any node \(i\) in “available vehicles” can be connected by any node \(h\) in “undersupply sub-regions” and the capacity is set as 1. The capacity of each edge from node \(h\) to node \(i\) is marked as \(D_h^m\). The objective function and constraints are shown in Equation (12)-(13). It is worth noting that the third decision variable \(x_{hj}\) is equal to \(f_{hj} (i \neq s, t\), and comes from available vehicles; \(j \neq s, t\), and comes from undersupply sub-regions).

\[
\text{Max} \ V = f^*
\]

(12)
Table 2. Flowchart of Kuhn-Munkres Algorithm for Charging or Delivery Task Assignment.

| Step | Description |
|------|-------------|
| Step 1: Initialize set \( S \) | If designing charging task assignment policy:  
Set \( S \) to \( S = S \cup \{v_1, v_2\} \).  
Else if designing delivery task assignment policy:  
Set \( S \) to \( S = S \cup \{v_1, v_2\} \). |
| Step 2: Find Alternating Path | Until there is no more vertices to add to the Alternating Path  
Step 3: Reset vertex values of \( V_1 \) and \( V_2 \)  
For \( i \in V_1, j \in V_2 \), \( w(i, j) = V_1^i + V_2^j = w(e)^0 \)  
Search for all the vertices both in subset \( V_1 \) and in the alternating path  
Search for all the vertices in subset \( V_2 \) while not in the alternating path  
Form a series of vertex pair \((i, j)\)  
Find the vertex pair whose \( d^+ \) is minimum  
The value of all the vertex in \( V_1 \) minus \( d^+ \), the value of all the vertex in \( V_2 \) plus \( d^+ \) |
| Step 4: Repeat Step 2 and Step 3 |  
Until finding the complete matching |
| Step 5: Output the final matching | Transfer the final matching result into the best task allocation instructions |

\[
\begin{align*}
\sum f_{ij} - \sum f_{ji} &= 0 (i \neq s, t) \\
\sum f_{ij} - \sum f_{ji} &= \kappa(t)(i = s) \\
\sum f_{ij} - \sum f_{ji} &= -v(f)(i = t) \\
f_{ij} &= (0, 1) (i \neq s; j \neq t)
\end{align*}
\]  

(13)

4. Designing policies

To design an optimal policy including charging, delivery, repositioning, and parking scheme for SAEVs fleet at each time period, the method framework for designing dispatching policies is as follows. First, KM (Kuhn-Munkres) algorithm is adopted to find the best charging task assignment policy for vehicles with low electricity. Second, the deep Q-learning algorithm is introduced to achieve a long-term reward \( r_{\text{mp}} \) considering a globally multi-step return to improve the order dispatching effect, and KM algorithm is then used to find the optimal delivery task assignment policy based on a new reward value. Third, Ek (Edmond Karp) algorithm is adopted to solve the above max flow problem and achieve the best vehicle repositioning scheme among different sub-regions.

In the bipartite graph \( G = (V, E) \), the KM algorithm is primarily used to determine the maximum weight matching under complete matching. With a temporal complexity of \( O(n^3) \), the KM method can reach an exact solution, making it an efficient and quick technique for real-large-scale applications. Assume a vertex \( i \in V_1 \) and a vertex \( j \in V_2 \) exists. \( \forall i \in V_1 \) and \( \forall j \in V_2 \), \( V_1^i + V_2^j \geq w(e)^0 \) (Wang and Guo, 2021).

For the charging task assignment problem,\( V_1 \) represents a vertex subset including SAEVs with the electricity of less than \( \Lambda \) kWh, \( V_2 \) represents a vertex subset including empty charging piles, \( w(e) \) means the weight between vertex \( i \in V_1 \) and \( j \in V_2 \) and is determined by \( r_{\text{mp}} \). For the delivery task assignment problem, \( V_1 \) represents a vertex subset including SAEVs with the electricity of more than \( \Lambda \) kWh, \( V_2 \) represents a vertex subset including trip orders waiting for pickup, \( w(e) \) means the weight between vertex \( i \in V_1 \) and \( j \in V_2 \), and it is determined by \( r_{\text{mp}} \).

Theorem 1. If vertex \( i \in V_1 \) and vertex \( j \in V_2 \) meet the condition \( V_1^i + V_2^j = w(e)^0 \), a set called \( S \) is defined and made up of directed lines from \( i \) to \( j \). If a match is the complete matching of the set \( S \), this match is also the maximum complete matching of the graph \( G \) (Wang and Guo, 2021).

According to Theorem 1, because the charging and delivery task assignment processes can be modeled as two conventional maximum weight matching problems, the KM algorithm is used to solve the final best matching results to assign SAEVs. The flow chart of KM algorithm is as follows in Table 2.

It is worth noting that in the execution code of KM algorithm, if the number of SAEVs and orders does not match, a new virtual SAEV or order (also known as a new virtual vertex) will generate automatically to realize the same number of SAEVs and orders. Although new virtual vertexes are added, the weighted value that starts from virtual vertex or ends at virtual vertex is set as 0. This way of handling first matches the principle of KM algorithm, and allows unfulfilled order requests at any time \( t \). For example, if two orders are matched with two virtual SAEVs, these two orders are unfulfilled, and will wait for the next round of order assignment.

Compared with delivery task assignment, the weight between each SAEV and each charging pile during the charging task assignment process is mainly related to \( r_{\text{mp}} \), which is linearly related to driving distance. Meanwhile, the least driving distance means the least energy consumption for these SAEVs with low electricity.

However, factors influencing delivery assignment results include order income and pickup duration (i.e. user waiting time). Order income is mainly related to driving distance and driving duration, which determines the economic income of the fleet operator. Pickup duration refers the period between the order initial time and user boarding time, which can effectively measure the user’s trip satisfaction. Besides, not only should the weight between each SAEV and user trip order consider an instant reward, but also a long-term return considering other orders in the next periods. Hence, \( r_{\text{mp}} \) is a complicated variable that involves multiple influencing factors.

To decide the final weight between each SAEV and each order, an instant reward calculated by the reward function (in Equation (22)) and an accumulated reward representing a long-term return calculated by the bellman function (in Equation (7)) are summed up as the final weight \( r_{\text{mp}} \). Based on previous fleet operating data, an experience trajectory should be constructed that includes information about SAEVs executing different tasks at different time periods. The \( V(s) \) value function will be modified repeatedly based on the experience trajectory using the Temporal Difference (TD) principle in Equation (14).

\[
V(s_t) \leftarrow V(s_t) + \alpha \left[ r_{t+1} + \gamma \cdot V(s_{t+1}) - V(s_t) \right] 
\]

(14)

To update the value function in previous experiments, the Q-learning algorithm was used. However, the Q-learning system can only store a limited number of past trajectories, but our country generates
billions of travel orders and accompanying trajectories every day. If the Q-learning method is going to be employed in large scale, a memory explosion will occur.

Therefore, a deep Q-learning algorithm, also known as a Back Propagation Deep Neural Network (BP-DNN) based estimator, is built to fit the value function V(s), as shown in Table 3.

Since vehicle repositioning task assignment can be modeled as a max flow problem, Edmond-Karp (EK) Algorithm is adopted to solve the max flow problem and the pseudo-code is illustrated in Table 4. Therefore, dispatching of SAEVs in several sub-regions including charging, delivery, and repositioning task assignment can be simultaneously completed at each time step based on KM algorithm, BP-DNN algorithm, and EK algorithm.

5. Numerical results

To validate the performance of the above combination of the combinatorial optimization method and deep reinforcement learning for solving the multi-task dynamic assigning problem of SAEVs, 4 simulation test cases are carried out by comparing the fleet operation effect under different conditions.

5.1. Dataset introduction

Four simulation test scenarios in this study include the real-time position of vehicles, Origin-Destination information of the daily trip order, location of the charging piles, and the sub-regions. The Didi platform’s open-source dataset is a one-month (November 1 to November 30, 2020) operational dataset of ride-hailing in Chengdu, Sichuan Province, China. It provides the vehicle and trip order data for this study. BAIC BJEV, a holding entity of Beijing Automotive Group Co., Ltd., has information regarding all the charging piles in Chengdu city. The dataset introduction is shown in Table 5.

Although the dataset used in this paper is from November 2020, the dataset still has general characteristics and is suitable for common situations. First, from the demand side, trip order data has characteristics of tidal, asymmetry, etc., which meets the actual daily trip distribution, and the attributes of trip orders such as actual travel distance, actual travel time, passenger waiting time, and actual order revenue have considered real-time traffic conditions. Second, from the supply side, the vehicle-to-pile ratio between the fleet size and the number of charging piles is about 3:1, which is the same as the vehicle-to-pile ratio in 2020-2022. Third, the dataset reflects the situation of insufficient supply in the fleet operation process in different regions and different time periods, and the dynamic dispatching demands for delivery, charging, and repositioning remain unchanged. Therefore, the dataset used in this paper can effectively verify the actual performance of fleet dispatching.

5.2. Description of test cases

In this paper, the baseline case only considers the income-preferred delivery reward function. To compare with the baseline case, 4 cases are designed to test the performance of the above methods. The first case mainly focuses on comparing the dispatching performance in 6 hours (from 8:00 am to 2:00 pm) under two kinds of delivery
reward functions (income preferred and income-satisfaction balanced) to validate the effectiveness of the new reward function $r_2$. In this case, charging task assignment and repositioning task assignment are not considered. The first case aims to compare the effect of 2 task assignment methods under different delivery reward functions.

The second case aims to compare the relocation performance in 6 hours (from 8:00 am to 2:00 pm) before and after considering charging task assignment based on the better reward function verified by the first case. In this case, repositioning task assignment among different sub-regions is not considered. The second case intends to figure out the benefit of considering delivery task with charging task together.

The third case centers on comparing the dispatching performance in 6 hours (from 8:00 am to 2:00 pm) before and after considering repositioning task assignment based on the better reward function verified by the first case. In this case, recharging task assignment, delivery task assignment, and repositioning task assignment are all considered at each time step, but the matching degree between each vehicle and delivery task is only represented by the instant reward function $r_2$. The third case aims to figure out the benefit of considering delivery task with charging task and repositioning task together.

The fourth case revolves around the improvement in 6 hours (from 8:00 am to 2:00 pm) before and after considering long-term return when assigning delivery tasks. The long-term return is represented by the state-value function $V(s, \theta)$ which is estimated by the BP-DNN estimator. In this case, recharging task assignment, delivery task assignment, and repositioning task assignment are all considered at each time step, and the matching degree between each vehicle and delivery task is represented by the sum of instant reward, and state-value $V(s')$. The fourth case aims to further improve the multi-task dispatching benefit by adopting a better reward calculation method.

5.3. Configuration of sub-regions, fleet, charging piles, and order requests

Before conducting 4 test cases based on the methodology put forward in this paper, some basic configuration is required, such as distribution of trip requests of each sub-regions, sub-regional information, and initialization configuration of the SAEV fleet. Fig. 3 depicts 19 hexagonal sub-regions of Chengdu’s inner ring in Sichuan Province, China, which are the areas to be investigated during the relocation process. The configuration data, which includes the amount of charging piles of each sub-region and the amount of SAEVs in each sub-region at 8 am is shown in Fig. 4. The radius of each sub-region (regular hexagon) is about 7 km considering that the inter-regional repositioning needs to be completed within 15 minutes (assume that the average driving speed of SAEV is 60 km/h).

Because vehicle recharging and delivery task assignment would be done independently in each sub-region, and the rebalancing task assignment should be done across many sub-regions, sub-region 1 was chosen as the target sub-area to show the performance of dispatching process. Sub-regions 1-7 are chosen as the target rebalancing areas in this study. The distribution of trip requests across 24 time steps (from 8:00 am-2:00 pm, set 15 minutes as a time step) across seven sub-regions is shown in Fig. 5 and Fig. 6.
5.4. Numerical results analysis

**Summary of test case 1:** During the dispatching process of SAEVs, changing the reward function from income-preferred to income-satisfaction balanced can realize obvious improvement in terms of user trip satisfaction and total order revenue.

Results of the first test case reveal that the reward function \( r_2 \) considering both order revenue and user trip satisfaction can realize a better operation performance when it comes to the waiting time of the passenger. As is shown in the frequency distribution in Fig. 7, the income-satisfaction-balanced situation shows obvious skew distribution compared with the income-preferred situation (the baseline situation in this experiment). Since the lower the average user waiting time is and the better the user trip satisfaction is, this new reward function \( r_2 \) can be referenced in the future SAEVs dispatching related research.

Test case 1 reveals that income-satisfaction balanced reward can help decrease user waiting time and then improve the user trip satisfaction. The income-satisfaction balanced reward is suggested to be adopted in the future SAEV dispatching process.

Meanwhile, the income-satisfaction-balanced situation can also bring a large improvement in accumulative order revenue. As is shown in Fig. 8, the accumulated revenue of the income-satisfaction-balanced situation in 6 hours is 33.2% more than the accumulated revenue of the income-preferred situation. This seems to be an interesting finding indicating that reducing user waiting time (pickup duration) can stimulate order revenue and improve user trip satisfaction at the same time. We believe that this is because less pickup duration can in reverse bring more delivery-on-road miles that should be billed.

**Summary of test case 2:** Introducing recharging task assignment into dynamic dispatching process of SAEVs can improve the density of vehicles at some time periods, and thus reduce the average pickup duration of some orders to improve user trip satisfaction. However, total order revenue will decrease due to the new-added recharging cost. De-
Fig. 7. Distribution of user waiting time of all the orders in test case 1.

Test case 1: delivery ✓ recharging × repositioning ×

Frequency distribution: represent the number of orders that fit in each subsection
Subsections: represent different passenger waiting time periods

Fig. 8. Revenue comparison per round of dispatch in test case 1 (24 time steps, unit:CNY).
Despite this, there is still a slight improvement for test case 2 than the test case 1 in terms of the total revenue.

Results of the second test case denoted that introducing recharging task assignment can increase the density of SAEVs during some time periods and thus reduce pickup duration to improve user trip satisfaction. As is shown in Fig. 9 and Fig. 10, there is a remarkable improvement in the aspect of user waiting time distribution, especially for orders whose ID ranks from 13000 to 25000 in the second half of total time period. The skewness becomes larger when taking the recharging task assignment into account. Specifically, this improvement mainly benefits from lower user waiting time of order 13000-25000. Since pile utilization rate decreases sharply and most of SAEVs are fully charged after time step 15, orders can be matched by SAEVs that are closer nearby. Hence, user waiting time of these orders can be reduced with the help of denser SAEVs.

However, it will pay a reasonable price for considering recharging task assignment in the aspect of order revenue. According to the assumption of 0.5 yuan per kWh, recharging cost will be a new-added expenditure for fleet operators. Fig. 11 reveals that there will be a 19.5% reduction in test case 2 compared with the situation without considering recharging in test case 1. Nevertheless, there is still a 7.3% increase compared with the situation adopting income-preferred reward function in test case 1. The above results proved the effectiveness of income-satisfaction balanced reward function indirectly.

**Summary of test case 3:** New consideration of repositioning task assignment in the dynamic dispatching process can realize great improvement in three aspects. First is timely vehicle replenishment can reduce user waiting time of completed orders to a great extent. Second is the total revenue considering repositioning task assignment can achieve a 50.0% improvement than test case 2. Third is repositioning can help SAEVs maintain a high order fulfillment rate during all the time steps.

Results of the third test case imply that introducing repositioning task assignment can improve user trip satisfaction to a large extent, but
Fig. 10. Pile utilization rate of 24 time steps of sub-region 1 in test case 2.

Fig. 11. Revenue comparison per round of dispatch in test case 1&2 (24 time steps, unit: CNY).

the availability of charging piles will be challenged. As illustrated in Fig. 12, when observing the passenger waiting time frequency distribution considering repositioning on the right, the skewness becomes larger than the situation without repositioning on the left. This indicates that pre-repositioning from oversupply sub-regions can increase fleet supply and thus average pickup duration of these trip orders can decrease. However, a larger vehicle supply volume also imposes more burdens on charging piles. Charging piles keep full capacity for 9 time steps (i.e. 135 min) during the 6-hour relocation process introducing repositioning task assignment, while the full capacity only lasts for 1 timestep (i.e. 15 min) during the 6-hour relocation process without repositioning task assignment. Hence, more charging piles should be built in advance in case of electricity shortage if repositioning task assignment is adopted when operating the SAEV fleet.

Also, the dispatching process with repositioning task assignment realizes a giant improvement in the total order revenue. Fig. 13 shows the revenue comparison of the above three test cases and it demonstrates that there is a 50.0% increase in the total revenue compared with the situation without repositioning task assignment. Revenue of the test case 3 ranks first among all the above 3 test cases mainly because repositioning can rebalance vehicle supply in different regions in advance so that more orders can be accomplished and higher order fulfillment ensures higher revenue.

Generally, results of the third test case show that both order revenue and user satisfaction can improve when taking account the repositioning task assignment. As is shown in Fig. 14, if the dispatching process only considers delivery, the recharging task without considering repositioning task assignment, the order fulfillment rate decreases sharply from 100% at the 10th time step to 2.8% at the 24th time step. In contrast, if the dispatching process takes repositioning task assignment into account, the order fulfillment rate can remain 100% among 24 time steps, which can satisfy user trip demand to the greatest extent. Meanwhile, order revenue among these 24 time steps will also increase by 50.0% from CNY 13598 to CNY 20122.

Summary of test case 4: Establishing a new reward function by adding the instant reward function with a BP-DNN value function representing long-term return can further achieve a slight improvement in terms of user trip satisfaction and total order revenue compared with the test case 3.

Results of the fourth test case show a slight improvement in the aspect of user trip satisfaction. Fig. 15 shows that the skewness of the frequency distribution in test case 4 becomes slightly larger. The num-
ber of completed orders whose pickup duration is less than 500 s is larger than the situation without considering a long-term return in test case 3, especially the increase of the orders with pickup duration of less than 200 s is more obvious. Besides, both test cases share a similar pile utilization rate. There is also a slight increase in terms of total order revenue. As is shown in Fig. 16, test case 4 adopting the accumulated reward function by adding the instant reward with a value function achieves higher revenue, with a 2.8% increase compared with the situation only considering the instant reward. Also, this outcome outperforms the revenues of all the other test cases.

Table 3 shows the training process of the loss function \((y - V(\phi_j, \theta))^2\) by using the Gradient Descent approach. More than 4000 completed historical order records are collected for training. Fig. 17 depicts the variation of loss value, with a consistent training loss of 4651.6009.

6. Conclusion and future work

This paper mainly aims to solve the multi-agent multi-task dynamic dispatching problem based on SAEVs in the future Autonomous Mobility-on-Demand system. The methodology proposed in this study achieves the best decision for allocating SAEVs to execute recharging, parking, delivery, and repositioning tasks.

To realize the above purpose, Markov Decision Process is adopted to model the decision-making process at first. Then dispatching process at each time step is broken down into trip order assignment, charging task assignment, and repositioning task assignment 3 sub-processes. Delivery and recharging task assignment are separately modeled as a maximum weight matching problem of the bipartite graph and solved by Kuhn-Munkres (KM) algorithm. Repositioning task assignment is quantified as a maximum flow problem and solved by Edmond-Karp algorithm. To further improve the performance of the above methodology, a new instant reward function balancing order income and user trip satisfaction
is newly designed, and a new accumulated reward function adding the instant reward with BP-DNN based state-value function is explored. To distinguish the effectiveness of the methodology, 4 simulation test cases are designed to verify the operational performance of the above methodology. Numerical results show some impressive findings:

1. Compared with the income-preferred reward function, the new income-satisfaction balanced instant reward function can help the total order revenue increase by 33.2% and decrease the average user waiting time (pickup duration) to a large extent.

2. Timely vehicle replenishment by repositioning task assignment in advance can help SAEVs in different sub-regions maintain a high order fulfillment rate during all the operating periods, thus resulting in a 50.0% increase in terms of total revenue and lower average passenger waiting time.

3. Summing the state-value function and instant reward function based on BP-DNN fitting, the new cumulative reward function can further achieve a small increase of 2.8% in total order revenue.

Thus far, numerical results of the above methodology prove that the combination of combinatorial optimization method and deep reinforcement learning can effectively solve multi-task dynamic dispatching problem of SAEVs in the future. In particular, considering recharging task assignment, delivery task assignment, and repositioning task assignment simultaneously at each time step and adopting accumulated reward function by adding instant reward function with BP-DNN based state-value function can significantly increase the total revenue and decrease the average passenger waiting time (pickup duration) at the same time. This methodology will be useful instruction for future SAEVs fleet operators.

We intend to improve this study in 4 areas in future projects. First, the final loss value of the state-value function that is BP-DNN based is still high and there is much room for reduction. More neural networks should be explored in the future to further decrease the loss value of the training dataset. Second, before conducting the repositioning task assignment, oversupply or undersupply sub-regions should be predetermined. This relies on a user trip request (trip demand) forecast model that can be compared with real-time vehicle supply volume. Third, to make recharging decisions for different SAEVs, an energy forecast model is also necessary to predict real-time remaining electricity of each SAEV. Therefore, a more accurate energy forecast model is also within the consideration for future research. Fourth, modeling
the repositioning task assignment process as a max flow problem can not ensure the least driving distance for the SAEVs with repositioning tasks. Converting the repositioning task assignment process into a min-cost flow problem is potential to overcome this limitation in the future.

**Declarations**

**Author contribution statement**

Ning Wang: Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data.

Jiahui Guo: Conceived and designed the experiments; Performed the experiments; Wrote the paper.

---

**Declaration of interests statement**

The authors declare no conflict of interest.

**Data availability statement**

The data that has been used is confidential.

**Funding statement**

Ning Wang was supported by Science and Technology Project of State Grid [5108-202119040A-0-0-00].

**Additional information**

No additional information is available for this paper.
Fig. 16. Revenue comparison per round of dispatch in test case 1&2&3&4 (time steps, unit:CNY).

Fig. 17. Variation in the loss value of historical episodes.

References

Aitouahmed, A., Josselin, D., Zhou, F., 2017. Relocation optimization of electric cars in one-way car-sharing systems: modeling, exact solving and heuristics algorithms. Int. J. Geogr. Inf. Sci. 32 (1), 1–32.

Al-Kanj, L., Nascimento, J., Powell, W.B., 2020. Approximate dynamic programming for planning a ride-hailing system using autonomous fleets of electric vehicles. Eur. J. Oper. Res. 284 (3), 1088–1106.

Alonso-Mora, J., Samaranayake, S., Wallar, A., Frazzoli, E., Rus, D., 2017. On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment. Proc. Natl. Acad. Sci. USA 114, 462–467.

Azevedo, C.L., Marcuz, K., Raveus, S., et al., 2016. Microsimulation of demand and supply of autonomous mobility on demand. Transp. Res. Rec. J. Transp. Res. Board 2564, 21–30.

Bahicheva, T., Burghout, W., Andreasson, L., Faul, N., 2018. The matching problem of empty vehicle redistribution in autonomous taxi systems. Proc. Comput. Sci. 130, 119–125.

Bauer, G.S., Greenblatt, J.B., Gerke, B.F., 2018. Cost, energy, and environmental impact of automated electric taxi fleets in Manhattan. Environ. Sci. Technol. 52 (8), 4920.

Boyac, B., Zografos, K.G., Gervoininis, N., 2015. An optimization framework for the development of efficient one-way car-sharing systems. Eur. J. Oper. Res. 240 (3), 718–735.

Brackers, K., Caris, A., Janssens, G.K., 2014. Exact and meta-heuristic approach for a general heterogeneous dial-a-ride problem with multiple depots. Transp. Res., Part B, Methodol. 67 (9), 166–186.

Chen, T.D., Kockelman, K.M., 2016. Management of a shared autonomous electric vehicle fleet: implications of pricing schemes. Transp. Res. Rec. J. Transp. Res. Board 2572, 37–46.

Dandl, F., Hyland, M., Bogenberger, K., et al., 2019. Evaluating the impact of spatio-temporal demand forecast aggregation on the operational performance of shared autonomous mobility fleets. Transportation 46 (6), 1–22.

De Almeida Correia, G.H., Antunes, A.P., 2012. Optimization approach to depot location and trip selection in one-way car-sharing systems. Transp. Res., Part B, Logist. Transp. Rev. 48 (1), 233–247.

Fagnant, D.J., Kockelman, K.M., 2014. The travel and environmental implications of shared autonomous vehicles, using agent-based model scenarios. Transp. Res., Part C, Emerg. Technol. 40, 1–13.

Fagnant, D.J., Kockelman, K.M., 2015. Dynamic ride-sharing and fleet sizing for a system of shared autonomous vehicles in Austin, Texas. Transportation 45 (1), 1–16.

Farhan, J., Chen, T.D., 2018. Impact of ridesharing on operational efficiency of shared autonomous electric vehicle fleet. Transp. Res., Part C, Emerg. Technol. 93 (8), 310–321.

Gurumurthy, K.M., Kockelman, K.M., 2018. Analyzing the dynamic ride-sharing potential for shared autonomous vehicle fleets using cellphone data from Orlando, Florida. Comput. Environ. Urban Syst. 71 (9), 177–185.

Hörl, S., 2017. Agent-based simulation of autonomous taxi services with dynamic demand response. Proc. Comput. Sci. 119, 899–904.

Hua, Y.K., Zhao, A.P., Wang, X., Li, X.P., 2019. Joint infrastructure planning and fleet management for one-way electric car sharing under time-varying uncertain demand. Transp. Res., Part B, Methodol. 128, 185–206.

Iacobucci, R., McElhanan, B., Terzika, T., 2019a. Optimization of shared autonomous electric vehicles operations with charge scheduling and vehicle-to-grid. Transp. Res., Part C, Emerg. Technol. 100 (3), 34–52.

Iacobucci, R., McElhanan, B., Terzika, T., 2019b. Costs and carbon emissions of shared autonomous electric vehicles in a Virtual Power Plant and Microgrid with renewable energy. Energy Proc. 156, 401–405.

Iglesias, R., Rossi, F., Zhang, R., Pavone, M., 2018. A BCMP network approach to modeling and controlling autonomous mobility-on-demand systems. Int. J. Robot. Res. 38 (2), 357–374.

Jüger, B., Agua, F.M.M., Lienskamp, M., 2017. Agent-based simulation of a shared, autonomous and electric on-demand mobility solution. In: IEEE International Conference on Intelligent Transportation Systems.

Jones, E.C., Leibowicz, B.D., 2019. Contributions of shared autonomous vehicles to climate change mitigation. Transp. Res., Part D, Transp. Environ. 72 (7), 279–298.

Jones, M., Feit, M.M., 2021. A generalization of Bellman’s equation with application to path planning, obstacle avoidance and invariant set estimation. Automatica 127, 109510.

Lammich, P., Seifgar, S.R., 2016. Formalizing the Edmonds-Karp algorithm. In: International Conference on Interactive Theorem Proving. Springer International Publishing.

Liang, X., Coelho, G.H.D.A., An, K., Areu, B.V., 2020. Automated taxis’ dial-a-ride problem with ride-sharing considering congestion-based dynamic travel times. Transp. Res., Part C, Emerg. Technol. 112, 260–281.

Mao, C., Liu, Y.L., Shen, Z.J., 2020. Dispatch of autonomous vehicles for taxi services: a deep reinforcement learning approach. Transp. Res., Part C, Emerg. Technol. 115, 102626.

Martínez, L.M., Viegas, J.M., 2017. Assessing the impacts of deploying a shared self-driving urban mobility system: an agent-based model applied to the city of Lisbon, Portugal. Int. J. Transp. Sci. Technol. 6 (1), 15–27.
Mendez, L.M., Bennassar, M.R., Chow, J.Y.J., 2017. Comparison of light rail streetcar against shared autonomous vehicle fleet for Brooklyn–Queens connector in New York city. Transp. Res. Rec. J. Transp. Res. Board 2650, 142–151.

Melendez, K.A., Das, T.K., Kwon, C., 2020. Optimal operation of a system of charging hubs and a fleet of shared autonomous electric vehicles. Appl. Energy 279, 115861.

Narayanan, S., Chamitakis, E., Antoniou, C., 2020. Shared autonomous vehicle services: a comprehensive review. Transp. Res., Part C, Emerg. Technol. 111, 255–293.

Nourinejad, M., Roorda, M.J., 2014. A dynamic carsharing decision support system. Transp. Res., Part E, Logist. Transp. Rev. 66, 36–50.

Nourinejad, M., Zhi, S., Bahrami, S., Roorda, M.J., 2015. Vehicle relocation and staff rebalancing in one-way carsharing systems. Transp. Res., Part E, Logist. Transp. Rev. 81, 98–113.

Paquette, J., Cordeau, J.-F., Laporte, G., Pascoal, M.M., 2013. Combining multicriteria analysis and tabu search for dial-a-ride problems. Transp. Res., Part B, Methodol. 52, 1–16.

Qin, W., Sun, Y.N., Zhuang, Z.L., Lu, Z.Y., Zhou, Y.M., 2021. Multi-agent reinforcement learning-based dynamic task assignment for vehicles in urban transportation system. Int. J. Prod. Econ. 240, 108251.

Rossi, F., Zhang, R., Hindy, Y., Pavone, M., 2016. Routing autonomous vehicles in congested transportation networks: structural properties and coordination algorithms. Auton. Robots.

Shaheen, S.A., Cohen, A.P., 2013. Carsharing and personal vehicle services: worldwide market developments and emerging trends. Int. J. Sustain. Transp. 7 (1–6), 5–34.

Shi, J., Gao, Y., Wang, W., Yu, N.P., Ioannou, P.A., 2019. Operating electric vehicle fleet for ride-hailing services with reinforcement learning. IEEE Trans. Intell. Transp. Syst. 99, 1–13.

Vosooghi, R., Puchinger, J., Jankovic, M., Vuillon, A., 2019. Shared autonomous vehicle simulation and service design. Transp. Res., Part C, Emerg. Technol. 107, 15–33.

Wang, N., Guo, J.H., Liu, X., Fang, T., 2019. A service demand forecasting model for one-way electric car-sharing systems combining long short-term memory networks with Granger causality test. J. Clean. Prod. 244, 1–16.

Wang, N., Guo, J.H., Liu, X., Liang, Y.Y., 2020. Electric vehicle carsharing optimization relocation model combining user relocation and staff relocation. Transp. Lett. 13 (4), 315–326.

Wang, N., Guo, J.H., 2021. Modeling and optimization of multi-vehicle dynamic dispatching problem for shared autonomous electric vehicles. J. Adv. Transp.

Xu, M., Meng, Q., Liu, Z., 2018a. Electric vehicle fleet size and trip pricing for one-way carsharing services considering vehicle relocation and personnel assignment. Transp. Res., Part B, Methodol. 111, 60–82.

Xu, Z., Li, Z., Guan, Q., et al., 2018b. Large-scale order dispatch in on-demand ride-hailing platforms: a learning and planning approach. In: The 24th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'18).

Zhang, W., Gubhathukurta, S., 2017. Parking spaces in the age of shared autonomous vehicles: how much parking will we need and where? Transp. Res. Rec. J. Transp. Res. Board 2651, 80–91.

Zhang, R., Pavone, M., 2014. Control of robotic mobility-on-demand systems: a queueing-theoretical perspective. Int. J. Robot. Res. 35 (1–3), 186–203.

Zhang, T.Z., Chen, T.D., 2019. Smart charging management for shared autonomous electric vehicle fleet: a Puget Sound case study. Transp. Res., Part D, Transp. Environ. 78, 102184.

Zheng, P., Wang, L.F., Zeng, Y.P., Chen, X., 2022. Back-propagation neural network modeling for a pulse tube refrigerator with passive displacer. Appl. Therm. Eng. 211, 118464.

Zhu, H.B., Liu, D.N., Zhang, S.Q., Zhu, Y., Teng, L.Y., Teng, S.H., 2016. Solving the many to many assignment problem by improving the Kuhn–Munkres algorithm with backtracking. Theor. Comput. Sci. 618, 30–41.