Neural Computing and Applications
https://doi.org/10.1007/s00521-024-09530-3

ORIGINAL ARTICLE

CASA: cost-effective EV charging scheduling based on deep reinforcement learning

Ao Zhang[1] · Qingzhi Liu[2] · Jinwei Liu[3] · Long Cheng[1]

Received: 21 June 2023 / Accepted: 14 January 2024

Abstract
With the widespread adoption of electric vehicles (EVs), the demand for public charging services is steadily increasing. Consequently, the development of effective charging scheduling strategies, aimed at optimizing the utilization of limited charging infrastructure, has become a key problem. Considering the diversity of user demands, we propose a Cost-Aware Charging Scheduling Architecture (CASA). This architecture considers both urgent and nonurgent charging customers by designing two charging modes with different power levels and associated costs. However, optimizing multiple objectives simultaneously while ensuring the interests of all parties involved in the charging demand response presents a challenge. Moreover, the uncertainty in customer charging demands and Time-of-Use (TOU) tariff further complicates the establishment of the model. To address the aforementioned challenges, this study formulates EV charging scheduling as a Markov Decision Process (MDP) based on deep reinforcement learning (DRL), employing the Deep Q-Network (DQN) algorithm for solution derivation. The objective is to minimize the operational costs of charging stations while ensuring the quality of service (QoS) requirements for customers. The simulation results demonstrate that CASA exhibits superior performance in optimizing both the average response time and service success rate, compared to commonly used baselines for charging scheduling. Furthermore, the CASA approach achieves a significant reduction in operating costs of EV charging station.

Keywords EV charging · DRL · Scheduling · Cost optimization

1 Introduction

In recent years, due to concerns about greenhouse gas emissions and the global warming effect, an increasing number of individuals have been inclined toward choosing environmentally friendly modes of transportation. Among these alternatives, electric vehicles (EVs) have emerged as a promising solution by utilizing cleaner energy sources to replace conventional fossil fuels, demonstrating significant potential in carbon reduction efforts. According to a report by the International Energy Agency [18], the global EV fleet is expected to reach 250 million vehicles by 2030, a substantial increase from the 5.1 million vehicles recorded in 2018. However, the widespread adoption of EVs has presented challenges, with one of the key issues being the escalation of peak loads and circuit overloads caused by uncoordinated charging. To address this, ensuring the stable and orderly integration of EVs into the grid has become a critical priority.

Public charging stations, designed to provide extensive charging services for local areas and are considered indispensable charging resources for EV users. Typically, these charging stations offer two distinct charging modes: alternating current Level 2 (AC II), which operates within a power range of 10–22 kW, and direct current fast charging

✉ Long Cheng
lcheng@ncepu.edu.cn

Qingzhi Liu
qingzhi.liu@wur.nl

Jinwei Liu
jinwei.liu@famu.edu

[1] School of Control and Computer Engineering, North China Electric Power University, Beijing, China

[2] Information Technology Group, Wageningen University and Research, Wageningen, The Netherlands

[3] Department of Computer and Information Sciences, Florida A &M University, Tallahassee, FL, USA

Published online: 21 February 2024

(DCFC), which provides a power range of 50–120 kW. In terms of electrical characteristics, AC II has longer charging times, while DCFC allows for shorter charging times but incurs higher power losses, which may accelerate battery aging [22]. In recent years, significant research has been conducted on charging station scheduling in various scenarios, including conventional charging stations [37], charging stations integrated with renewable energy sources [21], and distribution grid stations [27]. However, most of these studies assume that each charging station only provide a single charging service, either AC II or DCFC. This assumption falls short in accommodating the diverse service requirements of different charging users, such as shorter charging times, minimal battery degradation, or a combination of both. Relying solely on a single charging mode limits the ability to meet these varied needs, thereby impacting the overall service quality.

The implementation of a real-time and efficient charging scheduling strategy is crucial to ensure the smooth operation of charging stations, as it effectively mitigates potential risks associated with power system load fluctuations and transmission line overloads [1]. Unlike conventional refueling stations, EVs often experience significantly longer idle parking periods at charging stations, exceeding the time required for battery charging. Moreover, EV users are typically sensitive to the pricing of charging services [4]. They are willing to allow charging stations to accommodate the charging process based on the Time-of-Use (TOU) pricing policy, selecting lower tariff periods to minimize their electricity costs, provided that their charging needs are met without disruption. Furthermore, considering the stochastic and uncertain nature of customer behavior, if the allocation of charging station resources fails to align with the specific charging requirements of customers, it can lead to suboptimal charging efficiency and impact the quality of service (QoS) provided by the charging station. Specifically, the customer QoS requirements considered are declared upon their arrival at the charging station, with the goal is to satisfy these requirements by completing the charging within the expected time. Generally, the main focus of our study is to integrate unpredictable customer behavior and electricity price fluctuations, and we aim to optimize two vital areas: fulfilling customer QoS requirements and minimizing the operational costs of charging stations. This optimization is key to balancing the needs of both customers and service operators, ensuring customer satisfaction while promoting economic efficiency.

In response to customers' requests for charging services and to ensure the smooth operation of EV charging stations, extensive research has been conducted on the deployment of charging scheduling schemes. Some studies have attempted to integrate renewable energy sources into charging stations, aiming to achieve flexible energy scheduling and management schemes to further reduce energy costs [28]. Considering the impact of Time-of-Use policies on station operation, certain efforts have focused on developing dynamic pricing strategies to control the charging demand through economic incentives, thereby shifting the uncontrolled load [37]. While the aforementioned works have yielded effective optimizations from the perspective of charging station operations, insufficient attention has been paid to the limited number of charging points within the station. In reality, considering EV charging scheduling under the constraint of limited charging facilities is crucial as it directly affects the initial investment costs for station construction, which is an essential factor for operators to achieve profitable operations.

Several relevant studies have considered the limited availability of charging facilities and proposed corresponding optimization approaches [2, 10, 12, 19, 33]. However, in the works [2, 10, 19], to simplify the problem, the number of chargers is treated as the overall capacity constraint of the charging station. Consequently, these studies do not explicitly discuss the allocation of specific chargers to individual EVs. The approach proposed in [33] aims to reduce the service dropping rate of the charging station under the constraint of limited chargers, but it treats arriving EVs as a cluster, neglecting the individual demands. In the study [12], the mathematical definition of the charger constraint is explicitly provided, considering the limited number within the station. However, the assumption that customers will immediately leave if all chargers are occupied upon arrival weakens its practicality. To the best of our knowledge, there is a scarcity of published research addressing the issue of EV charging scheduling while considering the impact of TOU pricing and making specific charger allocation decisions for individual EVs.

Regarding the limitations of the aforementioned works, taking into account the limited number of charging facilities and the realistic scenario of dual-mode user demands, we propose a Cost-Aware Charging Scheduling Architecture (CASA), which is based on deep reinforcement learning (DRL) to optimize charging scheduling. The CASA approach tackles challenges like TOU pricing changes, unpredictable user behavior, and aligning chargers with service requests. In essence, CASA strives to create an optimal charging schedule for stations, balancing customer QoS needs with reduced operational costs.

Generally, the contributions of this work can be summarized as follows:

– We propose a cost-aware charging scheduling model for an EV charging station. In addition, we incorporate

a dual-mode charging service and the constraints of limited charging facilities into the charging scheduling model, while accounting the stochastic nature of future charging requests.
- We formulate the proposed charging scheduling model as a MDP and develop a DRL-based solution method to effectively optimize the operational costs of the charging station, encompassing the TOU charging cost, power loss overhead, and QoS penalties for service delays.
- Through meticulously designed simulation experiments under various sets of distinct variable conditions, we have substantiated the efficacy of the proposed CASA in comparison with other typical efficient scheduling methods, including reinforcement learning approaches. This superiority encompassing the optimization of scheduling objectives, improvement in learning efficiency, as well as the enhancement of stability and transferability.

The remainder of this paper is organized as follows. In Sect. 2, we introduce the related work. We present the system model and problem formulation in Sect. 3. In Sect. 4, we present the specific details of the proposed CASA procedure. In Sect. 5, the performance of our CASA procedure is evaluated, and a summary of the whole approach is given in Sect. 6.

2 Related work

In recent years, significant research efforts have been conducted to address the inherent stochastic nature of EV charging scheduling problems, encompass traditional scheduling strategies such as dynamic programming and day-ahead scheduling methods [34]. For instance, the study [32] proposed a two-stage dynamic programming strategy that utilizes short-term future predictions and long-term estimates based on historical data to reduce energy costs. However, the effectiveness of dynamic programming-based approaches heavily relies on the accuracy of load forecasting, making it challenging to obtain optimal charging strategies in practical scenarios. To mitigate these challenges, researchers have introduced day-ahead scheduling methods and employed robust or stochastic optimization techniques to formulate scheduling strategies that minimize the impact of uncertainty factors [5]. For example, the work [38] proposed a decision-making approach based on the information gap to optimize day-ahead scheduling for EV fleets and address the uncertainty associated with electricity prices. Additionally, robust optimization techniques have also been employed to provide optimal charging strategies. Although these methods demonstrate a certain effectiveness in day-ahead EV charging scheduling scenarios, they may not be suitable for real-time settings due to the high uncertainty associated with customer charging demands and TOU price variations.

Reinforcement learning (RL), as a prominent machine learning technique, provides a real-time framework for the EV charging scheduling in dynamic environments. It facilitates interactive learning by leveraging acquired rewards and environmental interactions, eliminating the need for prior knowledge of the underlying system. This characteristic makes RL an attractive choice in this context. For instance, the work [26] employs RL methods to optimize charging scheduling and pricing strategies, proposing a feature-based state-value function linear approximator to handle time-varying continuous states and action spaces. The proposed approach was validated using actual electricity price data, demonstrating significant cost reductions. Furthermore, recent advancements in DRL techniques have addressed limitations such as the applicability of RL solely to discrete action spaces and the curse of dimensionality. By employing deep neural networks to approximate the Q-table, these techniques have showcased exceptional performance across various scheduling scenarios in domains similar to ours. Examples include network control in the Internet of Things (IoT), network communication control [14] and edge technologies [15], as well as energy optimization in cloud data centers [30]. Building upon these advancements, an adversarial imitation reinforcement learning framework, referred to as AIRL, was introduced in [13]. It serves as a deep generative model that tackles suboptimality issues during the training of scheduling strategies in DRL approaches. Moreover, the work [11] focuses on enhancing the stability and convergence of the training process in large-scale multi-agent RL scenarios, while concurrently presenting novel cooperation schemes among agents of different types.

In Table 1, we summarize the optimization objectives of the CASA procedure and other relevant cost-aware methods. Currently, an increasing number of DRL studies in the energy field prioritizing cost-aware task scheduling, aiming to meet user demands while achieving resource efficiency during execution, thereby reducing energy costs. For instance, the work [6] proposes a real-time scheduling approach based on DRL to optimize the monetary costs associated with job execution in large-scale cloud environments. Similarly, the study [25] address the EV charging scheduling problem using a model-free DRL approach. They employ LSTM networks to extract electricity price features and determine the optimal charging strategy, thus achieving cost savings and alleviating anxiety. Furthermore, in work [29] extends the concept of comprehensive anxiety in DRL-based scheduling methods. This concept

Table 1 Comparison with some typical works in the current literature

References	Main objective				
	Cost	Demand	Utilization	Anxiety	QoS
Paraskevas et al. [19]	✔	✘	✘	✘	✘
Hao et al. [12]	✔	✔	✘	✘	✘
Manchella et al [16]	✔	✘	✔	✘	✘
Wan et al. [25] Yan et al. [29]	✔	✘	✘	✔	✘
CASA	✔	✘	✘	✘	✔

considers driver experience and charging preferences to determine the optimal charging sequence that strikes a balance between charging costs and driver anxiety. Moreover, a scheduling method was proposed in [16] that incorporates insertion cost-aware to dynamically match EVs within a fleet. By leveraging DRL, they efficiently dispatch idle vehicles to high-demand areas, thereby reducing vehicle costs and enhancing fleet utilization. However, in these cost-aware EV scheduling studies, there has been limited consideration given to more pertinent charging station characteristics, such as limited station allocation and the relationship between charging power and energy loss. Additionally, these studies have scarcely addressed the impact of charging station service type matching on QoS for customers.

Different from the aforementioned efforts, in this article, we propose a novel approach called CASA based on DRL to investigate the EV charging scheduling within dual-mode charging stations under the constraint of limited charging facilities. Our approach takes into consideration constraints that are associated with service types and average response time. The CASA method aims to obtain a globally optimal charging scheduling strategy, which reduces the average response time and decrease the operating costs, including energy loss and TOU electricity price.

3 The proposed CASA system

In this section, we detail the general system model adopted in our CASA framework, and then present the specific formulation for the cost optimization problem of charging operators. For better reference, the important notation is listed in Table 2.

3.1 System model for charging scheduling

In this paper, the framework of our proposed CASA procedure is shown in Fig. 1. The intelligent charging scenario considered in this study involves an EV charging station with limited facilities, accompanied by the deployment of CASA controllers. The controller serves as the central component within the entire framework, responsible for real-time decision-making regarding the allocation of incoming EVs to specific chargers within the station. The charging station comprises M DC chargers and N AC chargers, which provide services to EVs in a dual charging mode to cater to their diverse requirements. It is noteworthy that AC chargers offer advantages such as lower charging losses and reduced costs associated with battery lifespan compared to DC chargers [20]. However, a trade-off exists as AC chargers necessitate significantly longer charging durations.

When the EVs arrive at the station, they submit their charging information to the CASA controller, include parameters such as the emergency type, duration of network integration, and battery state of charge (SoC) requirements. Subsequently, based on this information and the real-time TOU pricing from the current power grid, a charging strategy π is formulated. In accordance with this strategy, the charging station allocates suitable chargers to the arriving EVs, then determines whether they should commence charging immediately or join a charging queue based on the chargers' operational status. Meanwhile, a QoS evaluation model is defined, which sets constraints for each arriving EV. These constraints must be given priority

Table 2 The used notation

Notation	Meaning
Cid_i	The id of the i-th EV
$CType_i$	The type of the i-th EV
$arrivalT_i$	The arrival time of the i-th EV
E_i	The charging demand by i-th EV
DDL_i	The QoS requirement by i-th EV
Pid_j	The id of the j-th charger
$PType_j$	The type of the j-th charger
$Vele_j$	Charging power for EVs
$VLoss_j$	Energy loss of j-th charger per time unit
T_i	The responsible time of the i-th EV
T_i^{char}	The charging time of the i-th EV
T_{ij}^{wait}	The wait time of the i-th EV

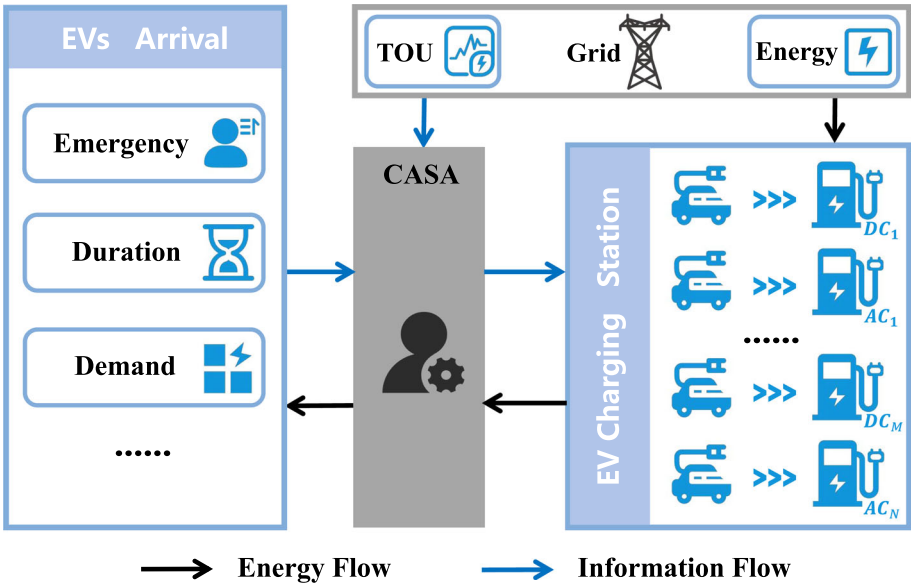

Fig. 1 The general system framework of CASA

fulfillment during the scheduling and execution phases. Furthermore, in order to model the optimization problems within this research, we provide mathematical definitions for the arrival of EVs and chargers, as well as other relevant definitions associated with the implementation of the scheduling mechanism in the CASA model.

Arrival EVs: an arrival EV $Car_i = \{Cid_i, CType_i, SoC_i, C_i, arrivalT_i, DDL_i\}$ is defined as the i-th charging request entering the charging station. Specifically, Cid_i is the EV id, which the controller allocates. $CType_i$ is a type identifier, which is used to classify EV charging requirements into two types: urgent charging demand and nonurgent charging demand. SoC_i is the battery state expected by the user for this charge request, which is a ratio of power level to achieve to complete this charge. Meanwhile, C_i is the battery capacity of this EV. $arrivalT_i$ is the arrival time of the EV, which the controller records. DDL_i is the estimated latest departure time of EV users, which is immediately submitted to the controller upon arrival at the charging station.

Chargers: in our study, we mainly consider two charging modes: DCFC and AC II. DCFC delivers high-voltage power directly to the EV battery through the charging port, resulting in fast charging speed but higher battery loss. On the other hand, AC II supplies alternating current to the EV onboard charger for battery charging, resulting in slower charging rates but lower battery degradation. As mentioned above, different customer types correspond to different service modes. Urgent customers prioritize reducing charging time, even if it leads to some battery lifespan degradation. Conversely, nonurgent customers prefer charging at a normal speed to minimize battery lifetime-related costs. By guiding EV users toward more suitable charging modes, it is possible to reduce congestion at charging stations while offering a greater variety of charging service options.

For a set of chargers in the charging station, the j-th charger can be represented as $CP_j = \{Pid_j, PType_j, Vele_j, VLoss_j\}$, where is the charger ID, the type of charger (i.e., DC type or AC type), the charging power for EVs, and the electrical loss rate of the charger, respectively.

TOU tariff: in our consideration, tariffs p_t play a significant role in the overall cost optimization problem. From the charger provider's perspective, the cost of electricity is determined by the price of purchasing electricity from the main Grid. To capture the realistic scenario of varying grid load severity, we divide the TOU tariff into three levels: peak period, normal period, and valley period [31]. We assume that the tariff for EV i is established at the moment of entering charger execution and remains constant throughout the charging process. Therefore, the CASA controller can learn to prioritize the lower valley period for allocating chargers to EV i to minimize the tariff cost while ensuring that the charging is completed within the given deadline, thus reducing the tariff cost and alleviating the load on the grid.

3.2 System optimization model

In the CASA approach, the priority for the charging provider is to develop an optimal scheduling strategy that effectively minimizes costs at each customer arrival point, by assigning each charging request of EV $i \in I$ to the most appropriate set of charger stations $J = [j_1, j_2, ..., j_n]$. In this case, the main optimization target is to minimize the total operating costs ω, which can be calculated as

$$\omega = \min \sum_{i=1}^{I} Cost_i \tag{1}$$

where $Cost_i$ represents the cost of EV i, specifically, consists of two main components, i.e., the cost of electricity purchased from the grid and the cost of power loss during charging. Therefore, in our optimization problem, the main focus is for the controller to learn how to make decisions that minimize $Cost_i$ at each time step t, thus achieving a reduction in the total overhead throughout the operation period, which yields

$$Cost_i = p_t * E_i + VLoss_j * T_i^{char} \tag{2}$$

In Eq. (2), the latter loss cost on charger j corresponding to EV i can be expressed as the product of the loss rate $VLoss_j$ and the charging time T_i^{char}, which will be presented next. And the former tariff cost is determined by the tariff p_t at the charging point t of EV i, and its charging demand E_i, which can be calculated as

$$E_i = SoC_i * C_i \tag{3}$$

$$E_{min} \leq E_i \leq E_{max} \tag{4}$$

The power required by the user for this charge, denoted as E_i, is determined by the product of the SoC_i submitted by the user and the capacity of the EV battery, denoted as C_i. Accordingly, we constrain E_i, to ensure it falls within certain limits. The maximum value, denoted as E_{max}, corresponds to the battery capacity C_i, representing the maximum power required when the battery can be fully charged. On the other hand, the minimum value, denoted as E_{min}, represents the minimum power required by the user.

The CASA method aims not only to minimize costs but also to meet customer QoS requirements by reducing the average response time. Our QoS evaluation model places significant emphasis on adhering to response time criteria, as this directly affects the success rate of charging services. The charging response time, represented as T_i, which encompasses the total duration between the submission of the charging request upon EV arrival and the completion of the charging process, two components should be taken into consideration.

$$T_i = T_i^{char} + T_{ij}^{wait} \tag{5}$$

Here, T_i^{char} indicates the time required for charging and T_{ij}^{wait} indicates the time spent in the waiting queue of EVs. Correspondingly, the charging time and queuing time can be defined as

$$T_i^{char} = E_i / Vele_j \tag{6}$$

where E_i indicates the power demand of EV i, while $Vele_j$ represent the charging power of chargers. Furthermore, assuming that there are $queueL_i^j$ charging requests waiting in the queue of the assigned charger j when EV i arrives, and EV i' is used to denote the n charging jobs assigned to charger j before EV i, then the waiting time for EV i can be calculated as

$$T_{ij}^{\text{wait}} = \begin{cases} \sum_{i'=0}^{n} T_{i'j}^{\text{char}}, & if \quad queueL_j^i > 0 \\ 0, & if \quad queueL_j^i = 0 \end{cases} \tag{7}$$

The given formula, inspired by Cheng et al. [7], indicates that the arriving EVs in the request queue are required to follow the first-come, first-served (FCFS) rule. Moreover, if a designated charger j is available, the charging request will be promptly addressed. By utilizing the above definition, we can introduce another crucial metric, denoted as $Resp_i$, which represents the reciprocal of the total response time for EV i. This establishes an inverse relationship between the two variables, serving as a critical measure for evaluating the charging station's timeliness in responding to each EV's charging request. It will subsequently be utilized in the computation of the reward function. This design guides the RL agent to make decisions that reduce the average response time, ensuring timely service completion and thereby upholding the guaranteed QoS for the customers.

$$Resp_i = \frac{1}{T_i} \tag{8}$$

As previously mentioned, EVs can arrive at the charging station at any time without a fixed pattern. For this scenario, it is essential that EVs are charged within the customer's expected time, as this greatly impacts service quality. Each charging request is linked to a deadline, or QoS requirement. A request is successful if charging finishes within this time, meeting the customer's QoS expectation. However, if charging exceeds the deadline, customers may end the process early and leave, leading to service failure. Following this approach and in accordance with the principles presented in [35], we outline a criterion for QoS success in CASA as:

$$\text{success}\left(\text{Car}_i, CP_j\right) = \begin{cases} 1, & \text{if } T_i \leq DDL_i \\ 0, & \text{else} \end{cases} \tag{9}$$

where DDL_i denotes the maximum acceptable response time specified upon the arrival of EV i. On this basis, the provided equation can be used to assess the success of assigning EV i to charger j in terms of meeting QoS requirements.

4 DRL-based implementation of the CASA procedure

In the proposed CASA procedure, we employ reinforcement learning to address the scheduling problem. RL is a machine learning technique, where an agent engages in continuous interactions with an unknown environment, taking specific actions to maximize the cumulative reward [23]. This process involves storing Q-function values in a lookup table, which can be challenging when dealing with high-dimensional environments. To overcome the curse of dimensionality, deep neural networks (DNNs) are introduced as a nonlinear Q value approximator, thereby defining the corresponding framework as a DRL method.

4.1 Preliminaries

When addressing problems using a DRL method, the mathematical model can be represented by a Markov Decision Process (MDP), taking the form of a quintuple (S, A, P, R, γ), where S denotes the set of all states that can be perceived in the environment, A denotes the set of all actions that the agent can take, P denotes the state transfer probability, R denotes the immediate reward under a particular state and action, and γ represents a value between 0 and 1, and serves to denote the significance of future rewards to the agent.

To solve the optimization problem presented in Sect. 3, we propose a DQN algorithm. The objective during the operation of EV charging stations over a duration of t time steps is to maximize the cumulative reward $R = \sum_0^t \gamma \cdot r_t$. As mentioned earlier, the discount factor γ is utilized to discount the reward. The DQN algorithm employs the optimal value $Q^*(s, a)$ to represent the maximum reward that can be obtained by taking action a_t in state s_t. The value $Q^*(s, a)$ can be iteratively obtained by solving the Bellman equation [3].

$$Q^*(s,a) = r_{t+1} + \gamma \max_{\pi} Q^*(s_{t+1}, a_{t+1}) \tag{10}$$

where s_{t+1} and a_{t+1} denote the state and action at the next moment; The Bellman equation shows that the value of the current state action is only related to the current reward and punishment value and the state-action value of the next step. During each iteration of the algorithm, the Q value undergoes updates as follows

$$\begin{aligned} Q_{t+1}(s_t, a_t) \leftarrow & Q(s_t, a_t) \\ & + \alpha\Big(r_t + \gamma \max_{a} Q(s_{t+1}, a_{t+1}) \\ & - Q(s_t, a_t)\Big) \end{aligned} \tag{11}$$

Given the parameter $\alpha \in (0, 1)$ as the learning rate, it signifies the extent of coverage over previous Q values. Moreover, to ensure that the agent is able to both explore the unknown environment and utilize the acquired knowledge, the selection of actions for the agent generally follows the ϵ-greedy strategy.

$$a = \begin{cases} \text{random}A & \beta < \varepsilon \\ \underset{a \in A}{\text{argmax}} Q(s,a) & \beta \geqslant \varepsilon \end{cases} \tag{12}$$

here, ϵ represents the fixed constant in the interval [0, 1]; β is randomly generated by the computer in the interval [0, 1]. When $\beta < \epsilon$, the agent randomly selects an action in the action space; otherwise, select the action with the greatest value in the current state.

Within the high-dimensional environment considered in our problem, the combination of all possible state-action values of EVs and charging stations forms huge state-action pairs. Consequently, the learning rate of the agent is excessively slow, rendering traditional RL agents (such as Q-learning) ineffective. DQN is proposed to overcome this issue. Unlike Q-learning, which employs a Q-table, DQN adopts a DNN to approximate the computation of Q values through a Q-network, as presented in Eq. (11). In the context of DRL, actions are not explicitly labeled or directly provided as training samples to the Q-network. Instead, the interaction between the agent and the unknown environment is leveraged to approximate this interaction. Currently, this DRL-based algorithm has proven to be effective and successful in various applications [24].

4.2 Markov decision process for CASA

In this section, we present the formulation of the EV charging scheduling model as a MDP with discrete-time steps. The primary goal is to determine an optimal strategy for the real-time scheduling of EVs and chargers, with the objective of minimizing the operational costs of the charging station. The specific details of each element within the MDP are presented as follows.

4.2.1 State space

We denote the state space as S, which consists of a set of vectors s_t at each time step t. The state s_t at a certain point t contains the necessary information about the arriving charging requests, as well as the state of the TOU tariffs and chargers in the environment, i.e., $s_t = \left(CType_i, E_i, DDL_i, p_t, T_{j1}^{wait}, T_{j1}^{wait}, ..., T_{jM}^{wait}\right)$, where $CType_i$ is the type of arriving EV, E_i and DDL_i are the charge demand and time requirements submitted by EV users. Besides, p_t represents the TOU tariff at step t, and T_{jM}^{wait} represents the waiting time on a charger at step t.

4.2.2 Action space

As previously mentioned, a separate request queue is established for each charge, and charging requests added to these queues are executed in a FCFS manner. Therefore, we define the action space as the set of all available chargers within the charging station, which are identified with a system-assigned ID represented as $a = (Pid_1, Pid_2, ..., Pid_M)$.

4.2.3 Reward function

Given any $s_t \in S$, $r(s_t, a_t, s_{t+1})$ is the immediate reward when a state transits from s_t to s_{t+1} over action a_t. Our objective is to minimize the operational costs of charging stations while concurrently ensuring the fulfillment of QoS requirements for customers. In a more detailed explanation, for each EV charging service provided, it is imperative to adhere to deadline constraints and, optimally, reduce the operational costs, which includes both charging costs and loss costs. As outlined in the problem formulation, the system should assign a charger to an arriving EV in a manner that ensures the charging process can be completed within the specified departure time while minimizing the associated cost. Therefore, the reward function should be designed to consider both the charging response time and cost. When the agent makes an assignment for EV i that results in a shorter response time, it should receive a higher reward. In terms of cost minimization, if EV i incurs a lower charging price and experiences fewer losses during charging, it should be rewarded. Otherwise, it will receive a penalty. In summary, the reward value r of EV i can be obtained through the reward function, which is expressed as

$$r = \left(1 + e^{\lambda - \mathrm{Cost}_i}\right)\mathrm{Resp}_i \tag{13}$$

where $Cost_i$ and $Resp_i$ are the main metrics we defined in the previous section, represents the charging cost and response time, respectively, λ is a hyper-parameter that enables a balanced control between cost and response time, which means that the immediate reward increases as charging time and power loss decrease.

4.3 CASA implementation

The optimal scheduling architecture for charging stations based on DQN is as follows. The overall workflow of the framework includes: Initially, the agent interacts with the environment to obtain transitions, which encompass states, actions, rewards, and next states. These transformations are then stored in a replay buffer, serving as a repository for past experiences. Finally, the agent samples a small batch of transitions from the buffer to update the neural network parameters. The following description delineates the specific workflow of the DRL-based CASA procedure and Algorithm 1 outlines this process.

Algorithm 1: The proposed CASA procedure

Implementation

Require: initial condition of EVs and Chargers, discount rate γ,exploration rate ε, replay memory capacity $\mathcal{C}$.

Initialize: mini-batch $S_\triangle$, reply period η, start learning time τ, target DNN with $\theta^* = \theta$, replay memory $\triangle$ to capacity $\mathcal{C}$;
for new EV i arrives at t **do**
  randomly select the value of probability μ;
  if $\mu \leq \varepsilon$ **then**
    randomly select EVs' actions a_t;
  else
    select actions that correspond to the largest Q-value estimated by evaluation,ie., $a_t = \arg\max_{a_t} Q(s_t, a_t; \theta)$;
  end if
  Schedule EV i according to action a_t, receive reward r_t, and observe state transition s_{t+1};
  store transition (s_t, a_t, r_t, s_{t+1}) in replay memory $\triangle$;
  if $j \geqslant \tau$ and $j \equiv 0 \bmod f$ **then**
    if $j \equiv 0 \bmod \eta$ **then**
      Reset $\hat{Q} = Q$;
    end if
  end if
  randomly select samples $S_\triangle$ from $\triangle$;
  update DNN parameters θ by performing a gradient descent on loss function;
  update parameters of target DNN after G iterations;
end for

Interaction: At the beginning of each step t, partial EVs arrive at the charging station and submit charging requests. By observing their service types, SoC requirements, departure times, and other relevant factors, the initial states of the EVs can be determined. Subsequently, the agent receives information regarding the waiting times at each charger within the charging station and real-time TOU prices from the main grid. By integrating this information, an observation $\{CType_i, E_i, DDL_i, p_t, T_{j1}^{wait}, T_{j1}^{wait}, ..., T_{jM}^{wait}\}$ is obtained. Then, the agent selects an action corresponding to the policy π derived from the network output. The agent continuously adjusts its scheduling strategy based on feedback provided by a defined reward function. This feedback guides the agent in improving the scheduling policy, aiming to enhance user QoS satisfaction while minimizing operational costs.

Experience replay: In this study, we encounter a challenge of learning an optimal charging scheduling strategy from scalar reward signals such as QoS and cost. This challenge is attributed to the sparse and delayed nature of rewards. Additionally, DRL-based approaches commonly assume a fixed distribution of underlying data, whereas real-world data often demonstrate partial correlations among sequences of samples. To address these issues, the utilization of an experience replay memory can improve the stability of the training process [17].

As depicted in Fig. 2, an experience replay buffer is established. At each time step t, after the agent interacts with the environment, the interaction experience can be stored as a quadruple (s_t, a_t, r_t, s_{t+1}), encompassing the current state, the taken action, the reward, and the subsequent state. Subsequently, during the training process, the agent no longer directly utilizes the current sample for training. Instead, it randomly samples a batch of records (referred to as a mini-batch) from the replay memory to update both the Q-network and the target network. Through this process, random sampling reduces the inter-sample correlations, while the replay mechanism mitigates noise during the training process, thereby enhancing the stability of the model.

Training: During the training process, the system first initializes its parameters and then proceeds to train the neural network extensively and effectively based on input data. Specifically, the DQN employs a dual-network architecture to enhance the convergence and stability of the training process. The Q-network aims to minimize the discrepancy between the current estimated Q values and the target Q values, allowing the agent to select optimal actions based on the current state. On the other hand, the target network is utilized to calculate the target Q values, serving as a reference during the training of the Q-network. These networks are responsible for action selection and evaluation, respectively. A loss function is employed to measure the difference between the predicted values and the ground truth, then updated the parameters using stochastic gradient descent to minimize this loss. To mitigate training fluctuations, the parameters of the Q-network are periodically copied to the target network, employing a mechanism known as delayed updates. Ultimately, by selecting the optimal parameters, the EV charging station is guided in choosing the most favorable action given a particular state.

To maintain effective exploration during the training of the scheduling policy, we employ an ε-greedy mechanism as the action selection strategy to choose the optimal actions. At the beginning of training, the agent can either

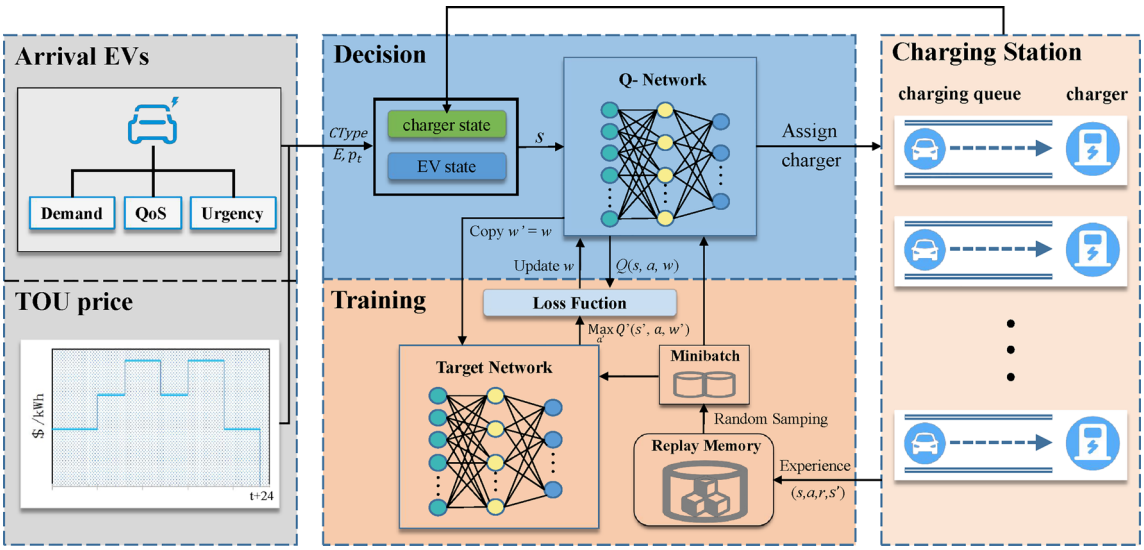

Fig. 2 The DRL-based architecture for EV charging scheduling

randomly select an action with a probability of ε to explore the action space or choose the action with the highest action value with a probability of 1-ε. Moreover, to strike a better balance between exploration and exploitation strategies, the exploration rate ε is decayed after each iteration at a decay rate until it reaches the minimum value ε_{min}. This approach encourages the agent to explore different actions at the beginning of training while avoiding excessive exploration in subsequent iterations.

5 Evaluation

This section entails a comparative analysis of the methodology we have proposed in contrast to several commonly employed charging scheduling techniques. We utilized Python 3.9 to construct our simulation environment, and our method is implemented within the Pytorch framework on a laptop equipped with an Intel Core i7-7700HQ CPU @ 2.80 GHz and 16GB of RAM. Through the experiments, our primary objective is to substantiate the superior performance of our CASA procedure concerning average response time and service success rate in comparison with extant scheduling methods, all while endeavoring to maintain cost-efficiency.

5.1 Experimental setup

We perform a series of evaluations simulating data to demonstrate the suitability of the CASA procedure for learning scheduling strategies in EV charging scenarios. Our evaluation considered a shared charging station with ten chargers, providing fast or slow charging services with capacities of 10, 20, 40, and 60 kW for EV charging. To capture their characteristics, we set loss rates for the chargers based on their respective charging capacities. For the evaluation, we utilized a typical TOU tariff schedule similar to Ding et al. [9], which is presented in Table 3. This schedule is characterized by hourly discrete segments, aligning well with the scale of the problem investigated in this experiment. Moreover, we model user behavior as a stochastic variable, where the arrival time, departure time, and power demand of EVs follow a truncated normal distribution. Specifically, taking charging demand as an example, we model these demands using a normal distribution, represented as $N \sim \mathcal{N}(20, 2^2)$, meaning that the charging demand of each arriving EV follows a normal distribution with a mean of 20 and a standard deviation of 4. This indicates that the range of charging demand exhibits a symmetrical probability distribution, with a major portion falling between 12 and 28 kWh, and the closer the demand is to 20 kWh, the higher the probability. Similarly, the departure time of EVs is determined by their estimated entry time, which follows a normal distribution $N \sim \mathcal{N}(0.5, 0.1)$. This implies that it predominantly falls within the range of $0.3-0.7$ hr. It is important to note that while we model the distribution of these random variables, our DRL-based approach does not rely on any knowledge of it, allowing the CASA procedure to be applied to different environments without requiring additional modeling.

Table 3 Tariff TOU Price (¥/kWh)

Time (h)	1	2	3	4	5	6	7	8
Tariff price	0.4	0.4	0.4	0.4	0.7	0.7	0.7	1
Time (h)	9	10	11	12	13	14	15	16
Tariff price	1	1	1	0.7	0.7	0.7	1	1
Time (h)	17	18	19	20	21	22	23	24
Tariff price	1	1	0.7	0.7	0.4	0.4	0.4	0.4

In our proposed CASA procedure, the underlying DNN was initialized using a feed-forward neural network architecture with a hidden layer containing 20 neurons. We set the replay memory value $N\Delta$ to 800 and the mini-batch size $S\Delta$ to 30. The learning rate was fixed at 0.01, and the target iterations were set to 50 decisions per episode. Following the general configuration for training the DQN model, we set the remaining parameters as follows: $\gamma = 0.9, f = 1, \tau = 500$, and ϵ is decreased from 0.9 by 0.002 in each learning iteration. In addition, the evaluation also include an implementation of the PPO (Proximal Policy Optimization) method to validate the performance of our proposed CASA based on the DQN. For specific parameters, we set the epsilon clipping $\epsilon - clip$ value to 0.2, and the entropy coefficient e to 0.01.

5.2 Experimental results

We evaluate several real-time methods for EV charging station scheduling, including two traditional approaches: the random method employs an unordered scheduling approach, where arriving EVs randomly select a charger. In contrast, the earliest assignment method employs a time-greedy strategy, assigning each arriving EV to the earliest available charging station to minimize the average response time. Additionally, we consider two DRL methods, in addition to the CASA method we proposed, which is based on the implementation of DQN, there is also a CASA reimplementation based on PPO as a baseline for evaluating the performance of RL agents. Both methods utilize networks of distinct architectures but achieve discrete action decisions, which are the two most widely applied approaches in similar scheduling problems [8, 36].

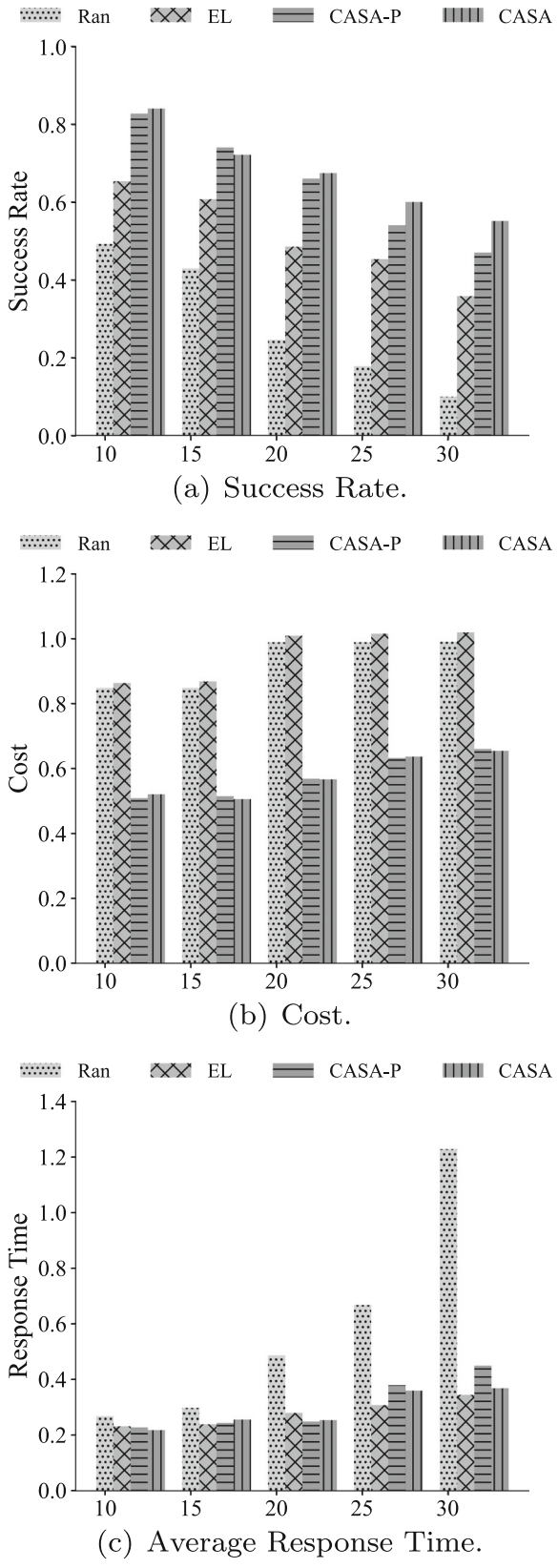

Fig. 3 Comparison by varying the mean EV arrival rate

It should be noted that in the subsequent experimental result figures, we use Ran to denote random selection, EL to denote earliest allocation, CASA-P and CASA, respectively, signify the implementations of our proposed method based on PPO and DQN algorithms.

5.2.1 Varying mean EV arrival rate

In this case study, we commence by comparing the performance of each method under a varying mean EV arrival rate. As mentioned earlier, the simulated charging station was configured with 10 chargers. To evaluate its scheduling capability, we intentionally set the number of EV arrivals per hour to exceed the number of available chargers at any given moment. Specifically, we varied the average EV arrival rate from 10 to 30, with a step size of 5. Furthermore, we maintained a balanced distribution of customer service types. The proportion of urgent and nonurgent EVs was uniformly set to 50%.

The results of this experiment are presented in Fig. 3. To avoid disproportionate influence on the agent's strategy learning due to unit differences in reward function, we normalized each metric. Therefore, in subsequent experimental results, "Cost" and "Average Response Time" are shown as abstract values without units. First, it is evident that our proposed CASA procure with two DRL implementations outperforms the other two baseline methods in terms of success rate, exhibiting an approximate 20% improvement compared to the suboptimal Earliest method. Additionally, as depicted in Fig. 3b, there is no significant cost difference among the first two baseline methods, while our CASA method effectively reduces costs by approximately 50%. Since the charging stations have a relatively fixed hourly response capacity under limited facilities, the cost of the charging station changes relatively little with the increase in EV arrival rates when viewed from the perspective of charging customers. Furthermore, Fig. 3c reveals that the CASA procedure achieves comparable performance to the earliest method in reducing the average response time, while the earliest method optimizes solely for minimizing response time in a greedy manner without regard for other metrics.

It is worth noting that both CASA-P and CASA demonstrate exceptional performance under low EV arrival rates. However, as the charging stations face a more intense service requests that exceed facility capacity, leading to performance fluctuations in the CASA procedure. Notably, when compared to CASA-P, CASA exhibits a more stable and gradual performance degradation, indicating superior adaptability. This phenomenon is contingent upon the distinct convergence and suitability of the two algorithms for the given problem scale, which will be further substantiated in subsequent experiments. In summary, the

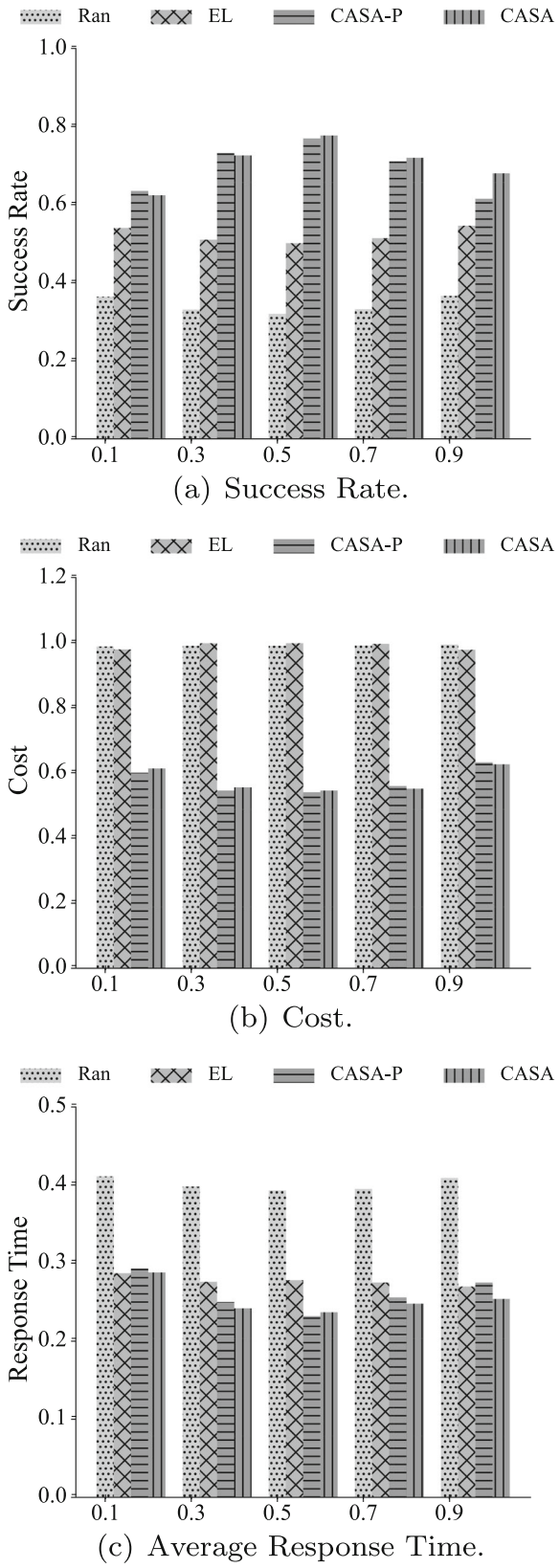

Fig. 4 Comparison by varying the proportion of EV types

CASA procedure we proposed exhibits the ability to reduce costs, shorten response times for real-time requests, and consistently outperform other baseline methods under various average EV arrival rates.

5.2.2 Varying urgent service proportion

In this experiment, we aimed to compare the performance of each method at different ratios of urgent service types. Unlike the previous case, we maintained a constant average EV arrival rate of 20 while varying the ratio of urgent service EVs from 10 to 90% in increments of 0.2. The ratio of AC II charger to DCFC charger in the charging station remained at 50%. The experimental results in Fig. 4 consistently demonstrate our method outperforming other baselines, including the PPO-based CASA-P, across all metrics. Our CASA procedure exhibits a success rate exceeding traditional comparative methods by approximately 30% to 50%, while maintaining a lower average response time compared to the earliest method. Furthermore, our method achieves a 40% cost reduction. Through the analysis of the individual subplots, it becomes evident that the CASA procedure performs exceptionally well when the proportion of urgent service customers is set to 50%. This is due to the ability of RL agent to make type matching decisions effectively, a capability that traditional methods clearly lack. Consequently, the first two baselines consistently exhibit inferior and relatively stable performance as the proportion of urgent service changes. In conclusion, our proposed CASA process exhibits superior performance at any service type ratio, and its performance is enhanced when the proportion of urgent and nonurgent customers is closer to parity.

5.2.3 Varying DCFC charger proportion

Similar to the case Sect. 5.2.2, in this experiment, we varied the charger types and evaluate the performance of our method accordingly. Here, a ratio of 0.1 indicates that only 10% of the chargers are DCFC chargers, while the remaining are AC II chargers. The experimental results, presented in Fig. 5, highlight a significant performance advantage of our method over the other baseline methods. In contrast to the other methods, which exhibit a degradation in performance as the charger proportions change, our method consistently maintains stable performance throughout the process. Particularly, when the proportions reach 0.7 and 0.9, the CASA procedure demonstrates significant performance advantages in terms of average response time and success rate. Furthermore, as depicted in Fig. 5b, CASA consistently achieves a stable cost reduction of approximately 40%. This experiment showcases that the CASA procedure can provide effective scheduling

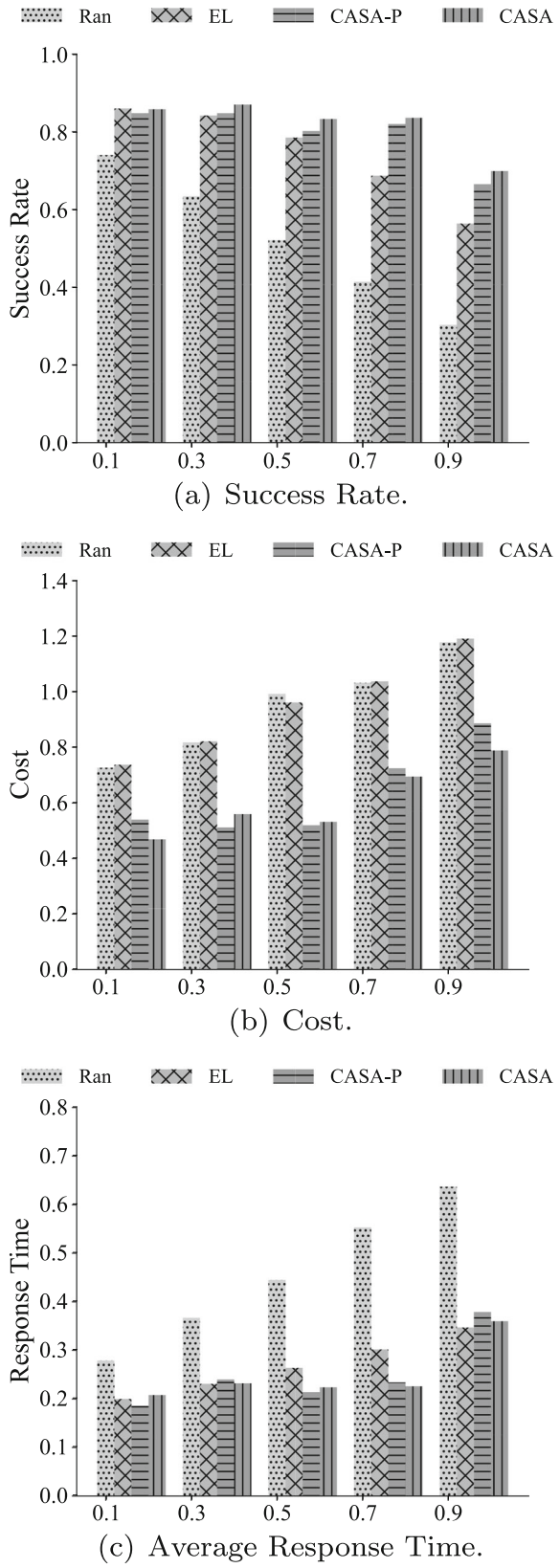

Fig. 5 Comparison by varying the proportion of charger types

regardless of the specific charger settings in the charging station, thereby demonstrating the scalability and adaptability of our proposed method.

The above three experiments provide insights into the performance of the CASA procedure in terms of reducing the average response time, improving the service success rate, and reducing the cost of EV charging stations. These experiments demonstrate the stability of our proposed method under different environmental conditions, as evidenced by the consistent performance across varying experimental variables. In summary, our DRL-based CASA procedure effectively reduces the overall cost of the charging process compared to traditional methods while maintaining user satisfaction and meeting service quality requirements. Moreover, it demonstrates the greater stability and improved performance of RL agent compared to CASA-P.

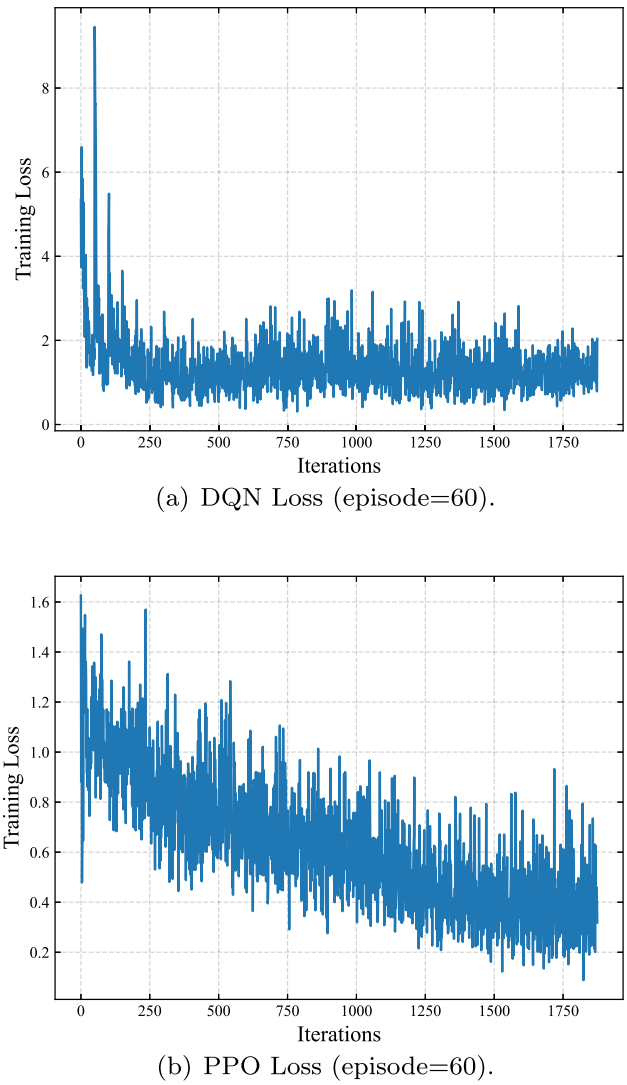

Fig. 6 Comparison of loss convergence between DQN and PPO

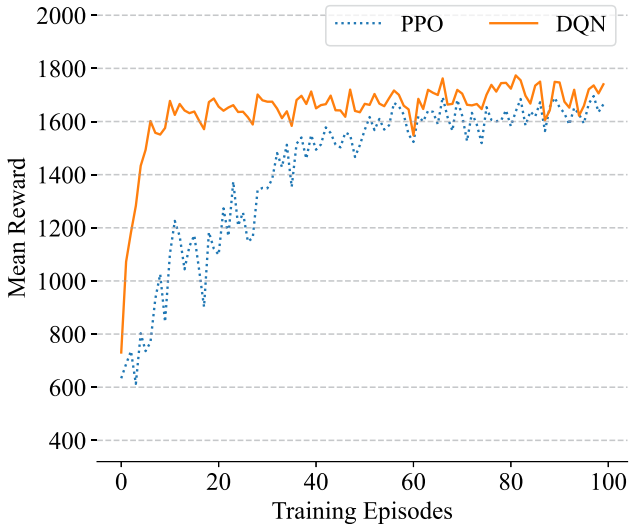

Fig. 7 Comparison of mean reward between DQN and PPO

5.2.4 Comparative analysis of DQN and PPO methods

In the aforementioned experiment, we assessed the performance of scheduling strategies optimized based on PPO and DQN by manipulating different variables. Preliminary findings indicate that our CASA implementation based on DQN exhibits superior convergence and stability. Subsequently, we will provide a comprehensive explication of this comparative outcome.

In the scheduling problem considered by CASA procure, where the action space consists of all the charger IDs within the charging station. The task of scheduling policy is to select the appropriate charger for a specific vehicle, making it a purely discrete decision problem. In this case, the stability of the Q-network structure employed by DQN is highly suitable. As illustrated in Fig. 6a, it calculates the loss through mean squared error to provide more stable estimates and can directly output the Q values of each charger for optimal selection. In contrast, as shown in Fig. 6b, PPO calculates the loss through both policy loss and value function loss, aiming to output a continuous action probability distribution or discrete action values. It is worth noting that due to the different loss calculation methods employed by the two algorithms, their loss ranges may different, while the convergence trends exhibited during the iteration process remain comparable. Therefore, DQN excels in this particular problem due to its specialization in handling purely discrete action decisions, without the need to accommodate the capacity for addressing continuous action spaces. In conclusion, the DRL scheduling strategy optimized based on DQN outperforms the re-implemented scheduling method using PPO in terms of reducing charging station costs and decreasing average response times. Furthermore, It is capable of meeting real-time scheduling demands for large-scale EV charging requests with continuous random arrivals.

For the comparison of transfer learning performance, PPO collects new data for online learning with each policy update, typically involving complex gradient policies. Conversely, DQN leverages the experience replay mechanism, sampling from previous experiences for parallel and iterative learning, which is generally easier to train. As depicted in Fig. 7, DQN exhibits convergence in average rewards after approximately 20 training episodes, whereas PPO requires approximately 60 episodes to achieve reward stability. Consequently, when the scale of facilities within the charging station needs to be adjusted, the more training-efficient DQN demonstrates superior transfer capabilities.

6 Conclusion

In this paper, we propose the CASA procedure, a cost-aware EV charging scheduling framework. The CASA framework considers the diverse charging demands of customers, addressing both urgent and nonurgent charging needs of EVs. To achieve this, we have designed two distinct charging modes with different power levels and corresponding costs for power losses, significantly improving responsiveness to user demands while reducing power costs. Furthermore, given the limited availability of charging infrastructure, the CASA procedure optimizes the allocation of chargers for each EV, aiming to minimize the charging cost based on the TOU tariff. We formulate the charging scheduling problem as a MDP and have developed an effective method using the DQN algorithm. Our experimental results demonstrate that the CASA procedure provides an efficient solution which meets QoS requirements and also significantly reduces the operational costs for the charging station, compared to existing methods.

Future work on EV charging scheduling can be further expanded with more detail. For instance, forthcoming studies could consider a broader range of real-world challenges, such as renewable energy generation, instead of solely relying on TOU tariffs from the grid to reflect the volatility of electrical loads. Furthermore, in future scheduling optimization endeavors, it would be pertinent to explore the impact of EV discharge behaviors and the heterogeneous characteristics of arrival models in different scenarios on charging strategies.

Acknowledgements This work was supported by supported by the Fundamental Research Funds for the Central Universities (2023YQ002).

Data availability Data will be made available on reasonable request.

Code availability Code is available at: https://github.com/zaNCEPU/CASA.

Declarations

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

References

1. Abdullah HM, Gastli A, Ben-Brahim L (2021) Reinforcement learning based EV charging management systems—a review. IEEE Access 9:41506–41531
2. Alinia B, Hajiesmaili MH, Crespi N (2019) Online EV charging scheduling with on-arrival commitment. IEEE Trans Intell Transp Syst 20(12):4524–4537
3. Barron E, Ishii H (1989) The Bellman equation for minimizing the maximum cost. Nonlinear Anal Theory Methods Appl 13(9):1067–1090
4. Chen Q, Wang F, Hodge BM, Zhang J, Li Z, Shafie-Khah M, Catalão JP (2017) Dynamic price vector formation model-based automatic demand response strategy for PV-assisted EV charging stations. IEEE Trans Smart Grid 8(6):2903–2915
5. Chen K, Ma Z, Zhou S, Shen X, Lin H (2020) Charging control strategy for electric vehicles based on two-stage multi-target optimization. Power Syst Prot Control 48:65–72
6. Cheng L, Kalapgar A, Jain A, Wang Y, Qin Y, Li Y, Liu C (2022) Cost-aware real-time job scheduling for hybrid cloud using deep reinforcement learning. Neural Comput Appl 34(21):18579–18593
7. Cheng L, Wang Y, Cheng F, Liu C, Zhao Z, Wang Y (2023) A deep reinforcement learning-based preemptive approach for cost-aware cloud job scheduling. IEEE Trans Sustain Comput. https://doi.org/10.1109/TSUSC.2023.3303898
8. Chen L, Yang F, Wu S, Xing Q (2021) Electric vehicle charging navigation strategy based on data driven and deep reinforcement learning. In: Proceedings of the 5th international conference on control engineering and artificial intelligence, pp 16–23
9. Ding T, Zeng Z, Bai J, Qin B, Yang Y, Shahidehpour M (2020) Optimal electric vehicle charging strategy with Markov decision process and reinforcement learning technique. IEEE Trans Ind Appl 56(5):5811–5823
10. Ghosh A, Aggarwal V (2017) Control of charging of electric vehicles through menu-based pricing. IEEE Trans Smart Grid 9(6):5918–5929
11. Guo J, Cheng L, Wang S (2023) CoTV: cooperative control for traffic light signals and connected autonomous vehicles using deep reinforcement learning. IEEE Trans Intell Transp Syst 24(10):10501–10512
12. Hao L, Jin J, Xu Y (2022) Laxity differentiated pricing and deadline differentiated threshold scheduling for a public electric vehicle charging station. IEEE Trans Ind Inf 18(9):6192–6202
13. Huang Y, Cheng L, Xue L, Liu C, Li Y, Li J, Ward T (2022) Deep adversarial imitation reinforcement learning for QoS-aware cloud job scheduling. IEEE Syst J 16(3):4232–4242
14. Liu Q, Cheng L, Jia AL, Liu C (2021) Deep reinforcement learning for communication flow control in wireless mesh networks. IEEE Netw 35(2):112–119
15. Liu Q, Cheng L, Ozcelebi T, Murphy J, Lukkien J (2019) Deep reinforcement learning for IoT network dynamic clustering in edge computing. In: IEEE/ACM international symposium on cluster, cloud and grid computing. IEEE, pp 600–603
16. Manchella K, Haliem M, Aggarwal V, Bhargava B (2021) Passgoodpool: joint passengers and goods fleet management with reinforcement learning aided pricing, matching, and route planning. IEEE Trans Intell Transp Syst 23(4):3866–3877
17. Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, Graves A, Riedmiller M, Fidjeland AK, Ostrovski G et al (2015) Human-level control through deep reinforcement learning. Nature 518(7540):529–533
18. Outlook IGE et al (2019) Scaling-up the transition to electric mobility. International Energy Agency, Paris
19. Paraskevas A, Aletras D, Chrysopoulos A, Marinopoulos A, Doukas DI (2022) Optimal management for EV charging stations: a win-win strategy for different stakeholders using constrained deep Q-learning. Energies 15(7):2323
20. Poullikkas A (2015) Sustainable options for electric vehicle technologies. Renew Sustain Energy Rev 41:1277–1287
21. Sun B, Huang Z, Tan X, Tsang DH (2016) Optimal scheduling for electric vehicle charging with discrete charging levels in distribution grid. IEEE Trans Smart Grid 9(2):624–634
22. Suresh P, Shobana S, Ramya G, Belsam Jeba Ananth MS (2023) Hybrid optimization enabled multi-aggregator-based charge scheduling of electric vehicle in internet of electric vehicles. Concurr Comput: Pract Exp 35(9):e7654
23. Sutton RS, Barto AG (2018) Reinforcement learning: an introduction. MIT Press
24. Torrado RR, Bontrager P, Togelius J, Liu J, Perez-Liebana D (2018) Deep reinforcement learning for general video game AI. In: 2018 IEEE conference on computational intelligence and games, pp 1–8
25. Wan Z, Li H, He H, Prokhorov D (2018) Model-free real-time EV charging scheduling based on deep reinforcement learning. IEEE Trans Smart Grid 10(5):5246–5257
26. Wang S, Bi S, Zhang YA (2019) Reinforcement learning for real-time pricing and scheduling control in EV charging stations. IEEE Trans Ind Inf 17(2):849–859
27. Wang J, Guo C, Yu C, Liang Y (2022) Virtual power plant containing electric vehicles scheduling strategies based on deep reinforcement learning. Electr Power Syst Res 205:107714
28. Yan Q, Zhang B, Kezunovic M (2018) Optimized operational cost reduction for an EV charging station integrated with battery energy storage and PV generation. IEEE Trans Smart Grid 10(2):2096–2106
29. Yan L, Chen X, Zhou J, Chen Y, Wen J (2021) Deep reinforcement learning for continuous electric vehicles charging control with dynamic user behaviors. IEEE Trans Smart Grid 12(6):5124–5134
30. Yan J, Huang Y, Gupta A, Gupta A, Liu C, Li J, Cheng L (2022) Energy-aware systems for real-time job scheduling in cloud data centers: A deep reinforcement learning approach. Comput Electr Eng 99:107688
31. Yang L, Dong C, Wan CJ, Ng CT (2013) Electricity time-of-use tariff with consumer behavior consideration. Int J Prod Econ 146(2):402–410
32. Zhang L, Li Y (2015) Optimal management for parking-lot electric vehicle charging by two-stage approximate dynamic programming. IEEE Trans Smart Grid 8(4):1722–1730
33. Zhang Y, You P, Cai L (2018) Optimal charging scheduling by pricing for EV charging station with dual charging modes. IEEE Trans Intell Transp Syst 20(9):3386–3396
34. Zhang C, Liu Y, Wu F, Tang B, Fan W (2020) Effective charging planning based on deep reinforcement learning for electric vehicles. IEEE Trans Intell Transp Syst 22(1):542–554
35. Zhang J, Cheng L, Liu C, Zhao Z, Mao Y (2023) Cost-aware scheduling systems for real-time workflows in cloud: An approach based on genetic algorithm and deep reinforcement learning. Expert Syst Appl 234:120972

36. Zhang Y, Chen X, Zhang Y (2022) Transfer deep reinforcement learning-based large-scale V2G continuous charging coordination with renewable energy sources. arXiv:2210.07013
37. Zhao Z, Lee CK (2022) Dynamic pricing for EV charging stations: a deep reinforcement learning approach. IEEE Trans Transp Electr 8(2):2456–2468
38. Zhao J, Wan C, Xu Z, Wang J (2015) Risk-based day-ahead scheduling of electric vehicle aggregator using information gap decision theory. IEEE Trans Smart Grid 8(4):1609–1618