Ad Hoc Networks 156 (2024) 103431

Contents lists available at ScienceDirect

Ad Hoc Networks

journal homepage: www.elsevier.com/locate/adhoc

A deep reinforcement learning approach for online mobile charging scheduling with optimal quality of sensing coverage in wireless rechargeable sensor networks

Jinglin Li [a,b], Haoran Wang [a,b], Chengpeng Jiang [a,b], Wendong Xiao [a,b,c,*]

[a] School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China
[b] Beijing Engineering Research Center of Industrial Spectrum Imaging, Beijing 100083, China
[c] Shunde Innovation School, University of Science and Technology Beijing, Guangdong 528399, China

ARTICLE INFO

Keywords:
Wireless rechargeable sensor network
Online mobile charging scheduling
Quality of sensing coverage
Deep Q-network
Multistage exploration strategy
Reward function

ABSTRACT

Mobile charging provides a new energy replenishment technology for Wireless Rechargeable Sensor Network (WRSN), where the Mobile Charger (MC) is employed for charging nodes sequentially according to the mobile charging scheduling result, using node charging timeliness and quality of sensing coverage as the scheduling criteria. Sensing coverage is a critical network property and has received more interest in recent research studies in mobile charging scheduling in WRSN. As the network environment is usually uncertain and the charging demands may change dynamically from time to time, online mobile charging scheduling is crucial, but existing online approaches are mostly based on specific network models, which are difficult to obtain in practical applications. In this paper, we propose a novel model-free deep reinforcement learning algorithm for the Online Mobile Charging Scheduling with optimal Quality of Sensing Coverage (OMCS-QSC) problem in WRSN, Multistage Exploration Deep Q-Network (MEDQN), where MC is designed as an agent to explore the online charging schedules via a new multistage exploration strategy for maximizing the network QSC according to the real-time network state. In addition, we also design a novel reward function to evaluate the MC charging action via the real-time sensing coverage contributions of the nodes. Extensive simulations show that MEDQN can reach the convergence state stably and is superior to existing online algorithms, especially in large-scale WRSNs.

1. Introduction

Wireless Sensor Network (WSN) is composed of a large number of sensor nodes spatially distributed for sensing the environment [1]. Due to its characteristics of low cost, scalability, self-organization dynamics, and fault tolerance, WSN is widely used in environmental monitoring [2], medical care [3], elderly care services [4], intelligent transportation [5] and manufacturing systems [6]. The limited network lifetime is the primary barrier to developing WSNs. To address this problem, researchers proposed the Wireless Rechargeable Sensor Network (WRSN), where the sensor nodes can be charged via wireless energy transfer.

Compared with the static charger in WRSN [7], using the Mobile Charger (MC) to charge the sensor nodes is more flexible and efficient [8]. Mobile charging scheduling is used to determine the charging sequence of the nodes according to the network demands. According to whether the MC charging sequence can be updated in time according to the real-time network state, existing mobile charging approaches can be divided to two categories: offline charging and online charging. In the offline approaches [9–16], before MC is ready to perform the charging task, the mobile charging sequence has been determined according to the known initial network state. However, the state and charging demands of the network may change due to uncertain factors, so the offline approaches are not applicable to this condition. The existing online approaches effectively solve this challenge [17–22], where MC can make charging decisions and adjust the charging sequence according to the real-time network state.

Sensing coverage is a critical network property and determines network credibility, and we evaluate the network coverage performance via the Quality of Sensing Coverage (QSC) in practical applications. When the network state is known and predictable, an offline mobile charging algorithm was proposed to maximize the network QSC by

* Corresponding author at: School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China.
E-mail addresses: B20190280@xs.ustb.edu.cn (J. Li), D202110337@xs.ustb.edu.cn (H. Wang), B20180299@xs.ustb.edu.cn (C. Jiang), wdxiao@ustb.edu.cn (W. Xiao).

https://doi.org/10.1016/j.adhoc.2024.103431
Received 18 July 2023; Received in revised form 23 January 2024; Accepted 31 January 2024
Available online 3 February 2024

optimizing the MC charging sequence [16]. However, the energy consumption rate of each node changes dynamically, and the network state (including the node's remaining energy and network coverage structure) is uncertain in dynamic WRSNs. If MC charges nodes still in a determined charging sequence, most nodes will not be charged in time and stop working, and the network QSC may be affected negatively. Therefore, this paper studies the problem of Online Mobile Charging Scheduling for optimal network QSC (OMCS-QSC), which aims to maximize network QSC in the charging cycle by finding the optimal mobile charging strategy according to the real-time network state. The difficulties of OMCS-QSC are as follows: (1) as the energy requirements of the sensor are dynamically changing under different charging time steps, the sensing coverage contribution of the same node and network charging demands cannot be accurately modeled; (2) all nodes cannot be scheduled for charging under the limited MC charging capability, how to select part of all nodes to charge to maximize the network QSC is difficult; (3) MC must reserve the energy for returning to the Charging Station (CS) under the limited battery capacity, OMCS-QSC can be modeled as a dynamic extended Traveling Salesman Problem (TSP) and is NP-complete in nature. Existing model-based online algorithms lack the impact of future charging action on network QSC in OMCS-QSC [23–25].

Reinforcement Learning (RL) is an essential branch of machine learning and has received more attention recently [26,27]. As RL can learn and improve the real-time action strategy through interaction with the environment under the reward-punishment mechanism, and considering the impact of future return on real-time actions while using the approximate estimation method during the learning process, RL has obvious advantages in real-time, dynamic and global optimal performance [28,29]. This paper will solve the OMCS-QSC problem based on RL. Since the representation of real-time network state in OMCS-QSC is complex, Q-learning as the basis of RL, is unsuitable for OMCS-QSC where the state–action function is fitted by a Q-table. Introducing the neural network to fit the state–action function of the complex network state, Deep Reinforcement Learning (DRL) was presented, and the effectiveness and efficiency of DRL have been demonstrated in dealing with the decision problems of complex state spaces in dynamic environments [30–33]. Therefore, based on the original Deep Q-Network algorithm of DRL, we present a novel model-free Multistage Exploration Deep Q-Network (MEDQN) algorithm for OMCS-QSC, where MC is taken as the agent to explore the space of the charging strategy iteratively via a new multistage exploration $\varepsilon_m - greedy$ strategy to maximize the network QSC according to the real-time network state.

The main contributions of this paper are summarized as follows:

1. Considering the effect of MC charging capability on network QSC in WRSNs, this paper studies the OMCS-QSC problem and proposes a novel MEDQN algorithm to maximize the network QSC by finding the optimal mobile charging strategy in the charging cycle.
2. In MEDQN, we design a new multistage exploration $\varepsilon_m - greedy$ strategy by introducing two strategy thresholds ε_m^1 and ε_m^2, making MC can select the current suboptimal actions with a certain probability during the exploration to improve the MC exploration efficiency.
3. A novel reward function with the charging penalty is designed according to the real-time sensing coverage contribution of each node to evaluate the MC charging action, which is defined as the weighted sum of the network loss sensing coverage ratio, the independent sensing coverage ratio and remaining energy ratio of the selected node.

The rest of the paper is organized as follows: the related works are introduced in Section 2. The system model and OMCS-QSC formulation are described in Section 3. In Section 4, the MEDQN algorithm for OMCS-QSC is presented. The performance of MEDQN is comparatively analyzed and discussed in Section 5. Conclusions are given in Section 6.

2. Related work

Existing studies on MC mobile charging scheduling in WRSNs mainly focus on MC charging performance and are divided to offline and online approaches. In the offline mobile charging approaches, MC charges nodes according to the definite charging sequence from the initial state information. For example, Wei et al. [9] proposed a multi-objective ant colony optimization algorithm to minimize average data transmission delay. To optimize the charging time of nodes, Jiang et al. [10] presented a quantum particle swarm optimization algorithm and defined a secondary performance index for charging waiting time. Srinivas et al. [11] proposed a hybrid optimization algorithm to reduce the moving distance and maximize mobile charging efficiency. Liu et al. [13] and Han et al. [14] presented a multi-node mobile charging scheme where MC can charge nodes within a specific charging range to reduce the number of off-working nodes. Xu et al. [15] determined each MC's independent closed charging path using multiple MCs in the multi-node mobile charging method to improve overall charging efficiency. Kan et al. [24] proposed an energy-recharging mechanism based on network connectivity to maximize the network quality of coverage. It initially constructed a mobile charging path by evaluating the nodes that sent charging requests outside their coverage and network connectivity contributions. In our previous work, the weighted sum of the network sensing-coverage ratio and the node survival ratio was taken as the new evaluation index, and we present an Improved Quantum Particle Swarm Optimization (IQPSO) mobile charging sequence algorithm to maximize the network QSC while considering both MC charging and moving time in [16].

The nodes' states change according to uncertain factors, MC updates the real-time mobile charging strategy according to the real-time network charging demands in the online mobile charging approaches. Researchers have proposed a variety of online algorithms with different mechanisms [17–20]. To maximize the network QSC in dynamic WRSNs, Yu et al. [23] and Dande et al. [34] presented a coverage-aware energy replenishment mechanism and a multi-node cost-effective charging scheduling algorithm, respectively. Real-time coverage contributions of each requested node and the benefits of chain-effect recharging coverage were all considered. However, the chain-effect recharging coverage benefit cannot be accurately estimated due to the dynamic nodes' energy consumption rates. Jiang et al. [25] considered the problem of on-demand scheduling MCs to maximize the covering utility, which quantifies the effectiveness of event monitoring, and three heuristics were proposed, but this scheme ignores the carrying capacity. As the network scale increases, existing heuristic algorithms make it difficult to find the optimal mobile charging decision. For example, in [20], some relatively important nodes with low energy consumption were often ignored, resulting in a long off-working time that reduced the network QSC. To overcome the problem that existing algorithms lack global optimality, researchers have introduced RL to generate the optimal mobile charging sequence in WRSNs.

RL is characterized by interacting with the environment in real time, where the agent continuously learns according to a reward feedback system that optimizes the action decision in the charging cycle [35]. Therefore, RL has shown great promise in the decision-making field of the Markov decision process [27]. For example, La et al. [29] designed a Q-learning algorithm to optimize the number of monitoring targets and the charging time. Wei et al. [27] and Soni et al. [28] introduced a charging path planning algorithm based on Q-learning that improves MC charging efficiency and extends the network lifetime. DRL is proposed with the neural networks to learn the multi-dimensional network state. Cao et al. [30] proposed a DRL charging algorithm that maximizes the sum of rewards collected by the MC under the constraint of MC capacity. Jiang et al. [31] and Yang et al. [32] proposed an actor-critic reinforcement learning algorithm to prolong the network lifetime while minimizing the number of non-working nodes.

Table 1
Comparison between literatures.

	Online scheduling	Sensing coverage	Survival of nodes	Capacity of MC
[27,36]	×	×	✓	×
[30]	✓	×	×	✓
[31,32]	✓	×	✓	✓
[24]	×	✓	×	✓
[16]	×	✓	✓	×
[25]	✓	✓	×	×
[23,34]	✓	✓	×	×
MEDQN	✓	✓	✓	✓

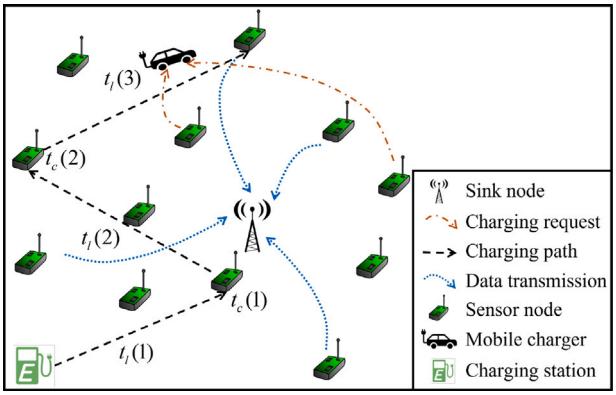

Fig. 1. The online mobile charging model of WRSNs.

We summarized the limitations of existing works in Table 1. Different from existing works, considering the impact of MC's charging capability (including charging power and battery capacity) on the network sensing coverage and node survival, we study the OMCS-QSC problem and propose a novel MEDQN algorithm.

3. System model and problem formulation

3.1. System model

As shown in Fig. 1, the WRSN studied in this paper consists of a Sink Node (SN) fixed in the center of the monitoring region, a Charging Station (CS), a MC and N rechargeable sensor nodes $S = \{s_1, s_2, \ldots, s_N\}$. The location $loc_{s_i} = (x_{s_i}, y_{s_i})$ of the node s_i is fixed and known, $i \in (1, N)$, and the location set of all nodes is $LOC = \{loc_{s_1}, loc_{s_2}, \ldots, loc_{s_N}\}$. SN can collect and process the real-time data from all nodes; MC can collect the real-time charging requests and the state information from all nodes, and then formulate the on-demand charging sequence according to the real-time network state, finally performing the charging task; CS is responsible for energizing MC. We assume the network deployment scenario is barrier-free and accessible. Table 2 summarizes the symbols used in this paper.

Definition 1 Charing Time Step (CTS): it is the step during which MC should be assigned a node for charging operation.

In this paper, we assume that each node has the same function but the energy consumption rate $V_{cs} = \{v_{cs}^1, \ldots, v_{cs}^N\}$ (J/s) and initial remaining energy $E_{in} = \{e_{in}^1, \ldots, e_{in}^N\}(kJ)$. r_p and $e_m(kJ)$ represent the charging request energy threshold percentage and the battery capacity of each node, when the real-time remaining energy of the node $e_r \le r_p e_m$, its charging request is sent to MC before the node runs out of energy. If $e_r = 0$, the node stops working and can resume working after being charged by MC. The real-time remaining energy of s_i in kth CTS is expressed as

$$e_r^i(k) = \begin{cases} e_r^i(k-1) - v_{cs}^i t_w(k) & l_i(k) = 0 \\ e_m & l_i(k) = 1 \end{cases} \quad (1)$$

Table 2
Symbolic descriptions.

Symbol	Description
R	Sensing coverage radius of nodes (m)
N	Number of nodes
L	Length of WRSN detection range (m)
C_E	Capacity of experience pool
D_{s_i}	Distance matrix between s_i and other sensors (m)
h_l	Number of hidden layers in the neural network
l_{h_l}	Number of neurons in the hidden layer
v_l	Travel energy consumption of MC (J/m)
d	Moving distance of MC (m)
e_{back}	Energy required for MC to return to CS (J)
a_I	Area independently covered by single node (m^2)
a_m	Maximum coverage area of each node (m^2)
a_{tm}	Maximum coverage area of WRSN (m^2)
a_l	Loss sensing coverage area of WRSN (m^2)
a_{TSA}	Total coverage area of all working nodes (m^2)
L_{MC}	The location of MC

where $l_i = 0$ means that s_i is not selected to charge and $l_i = 1$ indicates that it is scheduled for charging in the kth CTS. t_w is the total working time of MC, specifically expressed as $t_w(k) = t_l(k) + t_c(k)$ in Fig. 3, where MC moving time is $t_l(k) = d(k)/v_m$ and its charging time is $t_c(k) = (e_m - e_r^i(k))/v_r$, v_r stands for the received power of nodes from MC, $v_m(m/s)$ is the MC moving speed and $d(k)$ is the distance that MC moves to s_i. Since MC cannot ignore the transmission loss during wireless charging, we introduce the MC's power transmission efficiency θ_r, and $v_r = \theta_r v_c$, $v_c(J/s)$ is the MC's charging power.

We also assume that MC can receive charging requests continuously while performing the charging task; it can only charge one node at a time and leave when fully charged. v_c, v_m, θ_r and the consumption rate per unit moving distance $v_l(J/m)$ of MC are all unchanged in the charging cycle. Since this paper focuses more on the optimization of MC charging decisions, we assume that the distance between MC and nodes is close enough while MC is in the charging process, $\theta_r \approx 1$ and $v_r \approx v_c$.

Definition 2 Charging Cycle: charging cycle is the process from MC leaving CS to perform the charging task until returning to CS.

At the kth CTS, if MC chooses s_i, the real-time MC remaining energy $e_{MC}(J)$ is expressed as

$$e_{MC}(k) = e_{MC}(k-1) - (e_m - e_r^i(k-1)) - v_l d(k). \quad (2)$$

In WRSNs with randomly distributed nodes, incomplete or redundant sensing coverage will reduce the network QSC. We present a random node distribution mechanism to avoid the nodes' dense and scattered deployment, where the area covered jointly with three nodes will not be covered again by the fourth one. The specific process is shown in Fig. 2.

The distribution of nodes leads to a complex intersecting coverage of multiple nodes. Under the same CTS, selecting different nodes may cause different changes in the network coverage structure. The network Total Sensing-coverage Area (TSA) is the union of the sensing-coverage areas of all sensor nodes, and it can be calculated based on the real-time network state via the Monte Carlo method proposed in our previous work [37].

In OMCS-QSC, MC starts from CS and moves across the nodes that have sent charging requests in turn. Until the MC's remaining energy is insufficient, it returns to CS. If nodes cannot be charged by MC in time, the coverage vulnerability will occur in WRSN, resulting in incomplete network sensing information. In Fig. 3, taking four CTSs as an example, the initial energy set of five nodes is $E_{in} = \{10, 0, 20, 0, 30\}$, although $e_{in}^2 = e_{in}^4$, but it can be seen intuitively that the independent sensing area of the node $a_I^2 > a_I^4$, so the sensing coverage contribution of s_2 is greater than s_4, charging s_2 is the wise decision. Although s_4 stops working first, to ensure the optimal network sensing-coverage performance in the charging cycle, s_4 should be charged last.

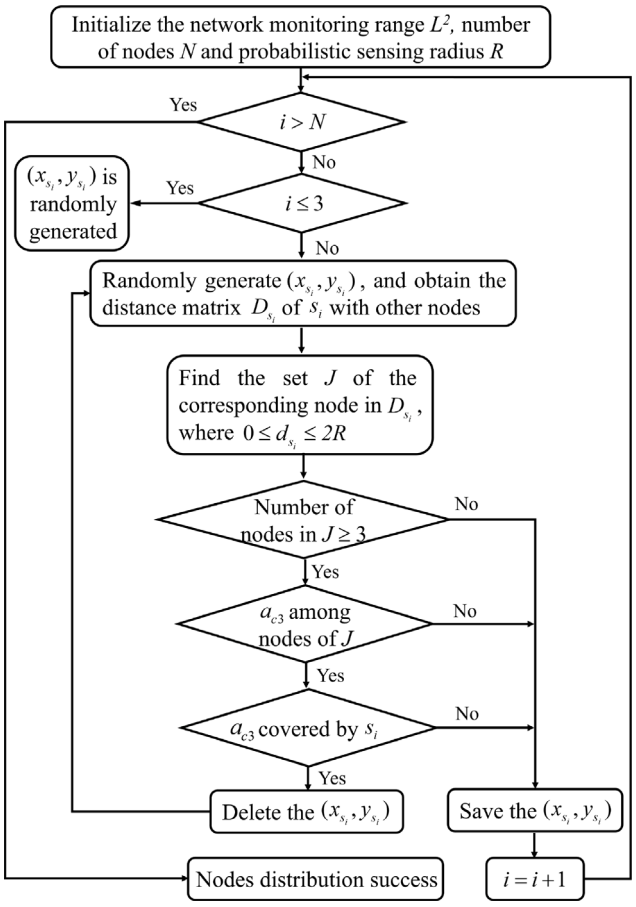

Fig. 2. Flow chart of the node distribution mechanism.

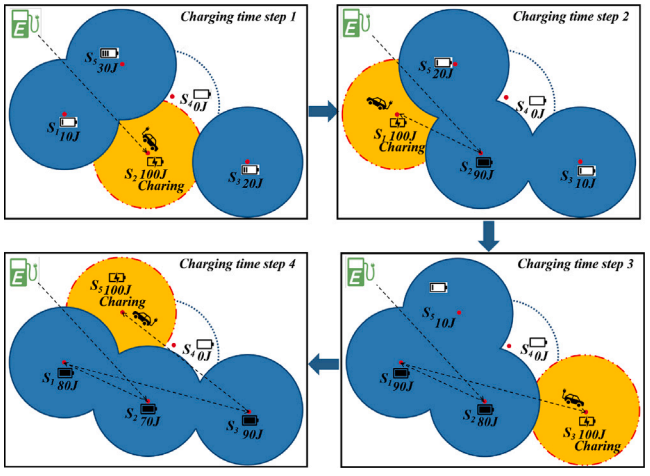

Fig. 3. Example of four CTSs in OMCS-QSC.

3.2. The formulation of OMCS-QSC

We formulate OMCS-QSC as a dynamic extended TSP, and it is also a nonlinear discrete variable optimization problem. The network QSC is evaluated by the weighted sum of the network Sensing-coverage Ratio and Node-survival Ratio (SRNR), expressed as $SRNR(k) = \delta_{pc}\frac{a_{TSA}(k)}{a_{tm}} + \delta_{wn}\frac{N-l_s(k)}{N}$, where δ_{pc} and δ_{wn} are the weight coefficients of network sensing-coverage and node survival, respectively, $\delta_{pc} + \delta_{wn} = 1$, $\delta_{pc} \in (0,1)$ and $\delta_{wn} \in (0,1)$; a_{tm} represents the maximum coverage area of WRSN, and the sensing-coverage ratio is the ratio of the real-time area covered by WRSN to a_{tm}; l_s is the number of off-working nodes. The average of the sum of SRNR for all CTSs in the charging cycle is defined as the network QSC, and maximizing the network QSC is the goal of OMCS-QSC. Therefore, the objective function of OMCS-QSC can be formalized as

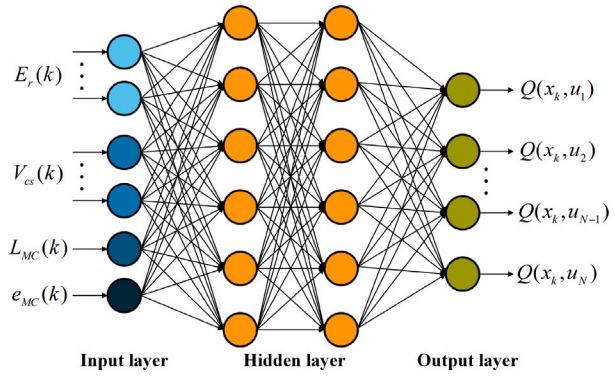

Fig. 4. Structure of the evaluation network N_{evl}.

$$
\begin{aligned}
Max \quad & QSC = \Big(\sum_{k=1}^{K} SRNR(k)\Big)/K \\
\text{s.t.} \quad & e_r^i(k-1) + t_c(k)v_r = e_m \\
& \sum_{k=1}^{K}\big(v_l d(k) + v_c t_c(k)\big) + e_{back}(k) \le E_m
\end{aligned} \tag{3}
$$

where K represents the total number of charging decisions in the charging cycle. The constraints of OMCS-QSC are as follows:

1. The sum of the remaining energy of nodes and the electricity supplemented by MC should be equal to the node's battery e_m.
2. The sum of energy supplemented to all nodes and moving energy cannot exceed maximum MC's battery capacity $E_m(kJ)$, the real-time MC's remaining energy should be greater than or equal to the real-time energy required by MC to return CS.

Different from the extended TSP in [16], MC may charge different nodes under the constraints of MC battery capacity in the charging cycle in OMCS-QSC. The particular case of OMCS-QSC has been proven to be NP-complete, where the MC battery capacity is large enough and the charging power of each node v_{cs} is unchanged in the charging cycle [16]. Therefore, it can be inferred that OMCS-QSC is also NP-complete.

4. The MEDQN algorithm for OMCS-QSC

4.1. The details of MEDQN

MEDQN consists of an environment state space X, an agent, and its action space U. At the kth CTS, the agent performs an action u_k on state x_k, which transfers to x_{k+1}, and the environment returns a reward r_k to the agent. The agent aims to learn an optimal mobile charging strategy to maximize total rewards. Under each state x, the state–action function value $Q_{evl}(x,u)$ is the return expectation of a charging action (charging a node) at the current WRSN state. While continuously exploring the environment, agents constantly learn and update Q_{evl}, which is represented by an evaluation neural network N_{evl} and shown in Fig. 4. To solve the OMCS-QSC problem, we set up MEDQN first.

1. Agent: it is the 'brain' of exploratory learning in the space of mobile charging strategies and is set to MC.

2. State: it is obtained by the agent by observing the environment and consists of the state of MC and network nodes. Considering the effects of the real-time remaining energy and energy consumption rates of nodes and the real-time location and remaining energy of MC, it is represented as $x = \{E_r, V_{cs}, L_{MC}, e_{MC}\}$.
3. Action: it is the charging action of MC, which represents the selection for charging nodes, $U = \{u|u = (1, 2, \dots, N + 1)\}$ is the action space, $u = N + 1$ is the action returning to CS and others represent the label of nodes in WRSN.
4. Reward: it is the immediate feedback to evaluate the actions of the agent. Since the reward is represented as the SRNR at each CTS, it lacks the penalty of the charging action, resulting in an infinite loop between several nodes with the larger independent sensing area a_I. Differently, considering the sensing coverage contribution of MC charging action and the impact of nodes' remaining energy on the network QSC, we design a novel reward function $r_k = a_I(k)/a_m - a_l(k)/a_{tm} - \beta e_r(k)/e_m$, where β is the reward penalty factor used to constrain MC charging decisions, $\beta > 1$, it means that the node's remaining energy has a stronger ability to constrain the MC charging decisions than the network real-time loss sensing coverage. Selecting the nodes with the lower remaining energy e_r, the loss sensing coverage area of WRSN a_l and the larger a_I is evaluated as more reasonable, and MC obtains a larger positive reward.
5. Strategy: it is the mapping of the relation between states and actions. Different from the traditional off-line $\varepsilon-greedy$ strategy, which makes MC only decide between the current optimal and random actions in the exploration phase, we designed an off-line multistage exploration $\varepsilon_m - greedy$ strategy for OMCS-QSC by introducing the other two thresholds ε_m^1 and ε_m^2, where MC can select the current suboptimal and sub-suboptimal actions with a certain probability during the exploration to improve the MC exploration performance. In the exploration phase, three initial thresholds are all large, and MC has a high probability of charging nodes randomly to achieve the purpose of exploration. When the exploration steps increase, ε_m gradually decreases, and nodes are selected with the best Q_{evl} value to obtain the maximum total rewards. The detailed $\varepsilon_m - greedy$ strategy is introduced in Algorithm 1.

Algorithm 1 $\varepsilon_m - greedy$ strategy for OMCS-QSC.

Require: N_{evl}, ε_m, ε_m^1, ε_m^2, $V_{cs}(k)$, $E_r(k)$, $L_{MC}(k)$ and $e_{MC}(k)$
Ensure: Real-time charging action $u(k)$

```
x_k = {E_r(k), V_cs(k), L_MC(k), e_MC(k)}
Q_evl(x_k, U) is the approximate return of all actions
Q_evl(x_k, U) = N_evl(x_k)
n = arg max(Q_evl(x_k, U))
U^1 is the action space where U removes n
n^1 = arg max(Q_evl(x_k, U^1))
U^2 is the action space where U^1 remove n^1 and n
n^2 = arg max(Q_evl(x_k, U^2))
Randomly generate a number l from 0 to 1
if l < ε_m then
    u_k = n^1
    if l < ε_m^1 then
        u_k = n^2
        if l < ε_m^2 then
            MC selects s_{n^0} to charge randomly
            u_k = n^0
        end if
    end if
else
    u_k = n
end if
Decrease ε_m as the iteration steps increases
```

MC aims to learn the optimal mobile charging strategy π^* to maximize the total rewards, then gets the real-time optimal mobile charging sequence Φ^*. We use temporal difference simulation and generalized policy iteration method to update the Q_{evl} function. First, estimate the Q_{evl} function value according to the given current action strategy; after obtaining the Q_{evl} value, update the action strategy in turn. At the kth CTS state transition of the gth iteration, when $x_k \in X$ and $u_k \in U$ are satisfied, the Q_{evl} function update is as follows

$$Q^{\pi}_{g+1,evl}(x_k, u_k) = (1 - \alpha)Q^{\pi}_{g,evl}(x_k, u_k) + \alpha(Q' - Q^{\pi}_{g,evl}(x_k, u_k)) \tag{4}$$

$$Q' = r_k + \gamma \max_{u_{k+1}} Q^{\pi}_{g,evl}(x_{k+1}, u_{k+1}) \tag{5}$$

$$\pi_{g+1}(x_{k+1}) = \arg\max_{u_{k+1}} Q^{\pi}_{g,evl}(x_{k+1}, u_{k+1}) \tag{6}$$

where α is the learning rate, γ is the return discount factor. As the real-time network demands may change dynamically, the energy consumption of nodes is uncertain. Therefore, MC may continuously explore the new network state, so the N_{evl} training results oscillate. MEDQN sets delayed updating to improve the stability of the training of N_{evl} by building a target function Q_{tar} via the target neural network N_{tar}. N_{evl} and N_{tar} have the same structure but play slightly different roles. N_{evl} is responsible for making charging decisions via the approximated $Q_{evl}(x, u)$; while N_{tar} is used to calculate the target value $Q_{tar}(x_k, u_k) = r_k + \gamma \max_{u_k} Q_{tar}(x_k, u_k)$ to stabilize the process of Q_{evl} iteration. By introducing $Q' = Q_{tar}$ into (4), we can obtain

$$Q^{\pi}_{g+1,evl}(x_k, u_k) = (1 - \alpha)Q^{\pi}_{g,evl}(x_k, u_k) + \alpha(Q_{tar}(x_k, u_k) - Q^{\pi}_{g,evl}(x_k, u_k)). \tag{7}$$

If we want to determine the optimal charging strategy $\pi^*(x_k)$, the optimal $Q^*_{evl}(x_k, u_k)$ must be obtained firstly. Therefore, the Q^*_{evl} with the convergent network parameter θ^*_{evl} satisfies $Q^*_{evl}(x_k, u_k) = r_k + \gamma \max_{u_{k+1}} Q^{\pi^*}_{evl}(x_{k+1}, u_{k+1})$ and $\pi^*(x_k) = \arg\max_{u_k} Q^*_{evl}(x_k, u_k)$. The specific parameter θ_{evl} of N_{evl} are updated as

$$\theta_{g+1,evl} \leftarrow \theta_{g,evl} - \alpha \nabla_{\theta_{g,evl}} (Q_{g,tar}(x_k, u_k) - Q_{g,evl}(x_k, u_k))^2 \tag{8}$$

Different from the update method of θ_{evl}, the target function parameter θ_{tar} is copied from θ_{evl} after a fixed number C iterations. The network state of adjacent CTSs is correlated. MEDQN builds an experience replay mechanism via a large-capacity C_E experience pool, which stores and updates the interaction state information of WRSN after each u_k. During the N_{evl} training process, MC randomly selects m groups of samples from the experience pool and can simultaneously learn from past and current experiences to address the problem of correlated states. As shown in Fig. 5, we show the structure of the MEDQN. In kth CTS, MC adopts the $\varepsilon_m - greedy$ charging strategy under the current network state x_k, executes u_k charge action and gets x_{k+1} and r_k. MC stores the interaction state information in the experience pool. During N_{evl} training, MC randomly selects samples to update parameters θ_{evl} via the experience replay mechanism. When the θ_{evl} error reaches the minimum training error ℓ_{evl} or the number of MC explorations reaches the maximum number of iterations G, the optimal charging strategy is obtained. The pseudocode of MEDQN is shown in Algorithm 2.

The theoretical convergence analysis and the derivation of MEDQN are described in [38], which we will not elaborate in this paper. To verify the convergence of MEDQN for OMCS-QSC, we compare the convergence performance of MEDQN and original DQN via the Matlab2021b simulation software. For the two network structures of MEDQN and DQN, we used a fully connected feedforward neural

Algorithm 2 MEDQN for OMCS-QSC.

Require: Initialized N_{evl}, N_{tar}, θ_{evl}, θ_{tar}, G, LOC, α, γ, ε_m, π, m, C, $V_{cs}(k)$, $E_r(k)$, $L_{MC}(k)$, $e_{MC}(k)$
Ensure: Well trained Q^*_{evl} and π^*

for $g = 1 : G$ **do**
  k=1
  while $k \geq 0$ **do**
    Obtain u_k from Algorithm. 1
    x_{k+1}, r_k, $e_{MC}(k)$ and $e_{back}(k)$ are generated
    Store $\{x_k, u_k, x_{k+1}, r_k\}$ into the experience pool
    Randomly sample m group experience
    if $e_{MC}(k) \leq e_{back}(k)$ **then**
      $Q^{\pi}_{evl}(x_k, u_k) = r_k$
      Train N_{evl} to update θ_{evl}
      Break
    else
      $Q_{evl}(x_k, u_k) = r_k + \gamma \max_{u_{k+1}} Q^{\pi}_{tar}(x_{k+1}, u_{k+1})$
      Train N_{evl} to update θ_{evl}
    end if
    if $g = C$ **then**
      $N_{tar} = N_{evl}$
    end if
    $x_k = x_{k+1}$
    $L_{MC}(k) = loc_{u_k}$
    $E_r(k) = E_r(k+1)$
    $k = k + 1$
  end while
  if The error of training θ_{evl} drops to ℓ_{evl} **then**
    Break
  end if
end for
We can obtain the optimal Q^*_{evl} and π^*

network with h_l hidden layers, which contain l_{h_l} neurons. As shown in Fig. 6, under the initialization parameters, the total rewards of MEDQN can converge from −61.23 to about 14.5 in about 3000 iteration steps, which verifies the convergence of MEDQN. The total reward obtained by the original DQN can converge from −32.84 to about 8 before 3000 iterations, and its convergence trend is relatively smooth. In contrast, MEDQN has better performance for OMCS-QSC and can make better-charging decisions than the original DQN to optimize the network QSC, which demonstrates that making changes to the exploration strategy in MEDQN is effective.

To verify the superiority of MEDQN compared with offline algorithms in responding to dynamic changes in network state, we compare MEDQN with the offline IQPSO algorithm [16] in terms of network QSC under different MC charging powers. As shown in Fig. 7, the network QSC obtained by the offline algorithm is always lower than that of the online MEDQN and original DQN algorithms under different MC charging powers, where MEDQN is superior to others. The reason is that the charging sequences generated by the offline algorithm based on the initial network state and cannot be adjusted in time according to changes in the network, so the offline algorithm is unsuitable for OMCS-QSC, and the superiority of MEDQN is demonstrated.

4.2. Time complexity

We calculate the time complexity of MEDQN from two aspects: environment interaction and training iteration.

In environment interaction, MEDQN needs to make charging decisions for K CTSs, observe the network's state transition and collect rewards. In addition, MEDQN can store the K exploration samples into the experience pool for training. Therefore, the time complexity of interacting with the environment is usually $O(K)$.

Table 3
Network environment simulation parameters.

Parameters	Value	Parameters	Value
L	100	v_{cs}	0.2 ∼ 1.8
R	6	δ_{pc}	0.8
N	50	e_m	144
E_{in}	43.2 ∼ 144	δ_{wn}	0.2
G	5000	v_m	5
α	0.5	E_m	72 000
v_c	40	ε_m	0.999
ε^1_m	$0.9\varepsilon_m$	ε^2_m	$0.8\varepsilon_m$
γ	0.97	v_l	10
β	2.2	r_p	0.3
m	1000	C_E	5000
ℓ_{evl}	0.00001	C	50
h_l	2	l_{h_l}	$2N + 2$

Table 4
Comparison of different α.

	0.1	0.2	0.3	0.4	0.5
QSC	0.893	0.891	0.892	0.891	**0.916**
TRs	−0.512	−5.138	−4.794	−3.945	**10.02**
SCR	0.899	0.896	0.895	0.894	**0.924**
NONs	6.696	6.386	5.886	6.179	**5.512**
	0.6	0.7	0.8	0.9	
QSC	0.914	0.904	0.912	0.898	
TRs	8.251	8.065	6.212	3.563	
SCR	0.921	0.911	0.919	0.904	
NONs	5.799	6.037	5.845	6.114	

In the training iteration of MEDQN, the evaluation N_{evl} needs to train G times and randomly select m samples from the experience pool. As N_{evl} has h_l hidden layers and each layer has l_{h_l} neurons, the time complexity of N_{evl} training is $O(Gh_l l_{h_l} m)$. The parameter θ_{tar} of the target N_{tar} is copied from N_{evl} after C iterations, the time complexity of N_{tar} updating is $O(G/C)$. Therefore, the time complexity of the training iteration is usually $O(Gh_l l_{h_l} m + (Gh_l/C))$. In summary, the total time complexity of MEDQN can be estimated as $O(Gh_l l_{h_l} m + (Gh_l/C) + K)$.

4.3. Parameter adjustment

Parameters of MEDQN will directly affect the algorithm performance and stability, so their adjustments are essential for OMCS-QSC. This section will use the comparison method to find the most suitable learning rate, return discount factor, and reward penalty factor for OMCS-QSC, respectively.

4.3.1. Learning rate

Learning rate α is a critical parameter that affects the learning step size and convergence stability of MEDQN. If the value of α is unchanged, MEDQN will fluctuate violently and converge slowly during training. Therefore, we will find the optimal α in this part for OMCS-QSC. The specific reduction strategy of α is $\alpha_g = \alpha_{g-1} - v_\alpha(\alpha_{g-1}/\alpha)$, where v_α is the descent rate. We compared the network QSC and other indexes (Total Rewards (TRs), Sensing Coverage Ratio (SCR) and the Number of Off-working Nodes (NONs)) of different α in the same network environment (shown in Table 3).

Table 4 shows MC has the best learning effect for the optimal charging strategy when $\alpha = 0.5$, the TRs is about 10.02, and the network QSC is about 0.9155. When $\alpha < 0.5$, TRs gradually decreases as α decreases. Because α is small, MC's exploration is incomplete within G, and the optimal mobile charging strategy cannot be learned. When $\alpha > 0.55$, the continual increase of α accelerates the learning speed. However, there are large fluctuations when the TRs converge to about 8. For example, for $\alpha = 0.9$, the last learning rate of MC's exploration α_{5000} is still around 0.18, and the learning step size is too large to achieve stable convergence at the end of the exploring.

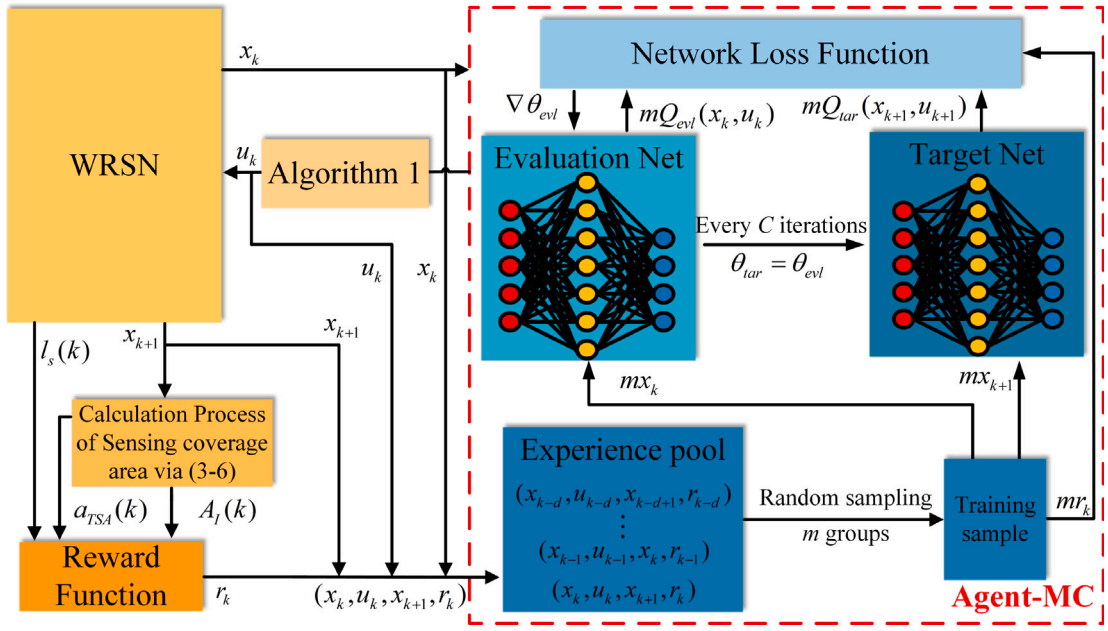

Fig. 5. The MEDQN structure for OMCS-QSC.

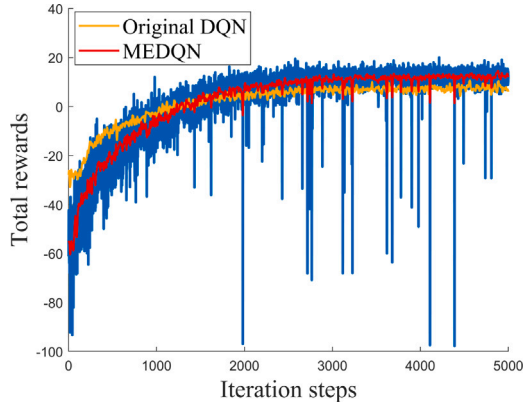

Fig. 6. Comparison of iterative convergence between MEDQN and original DQN.

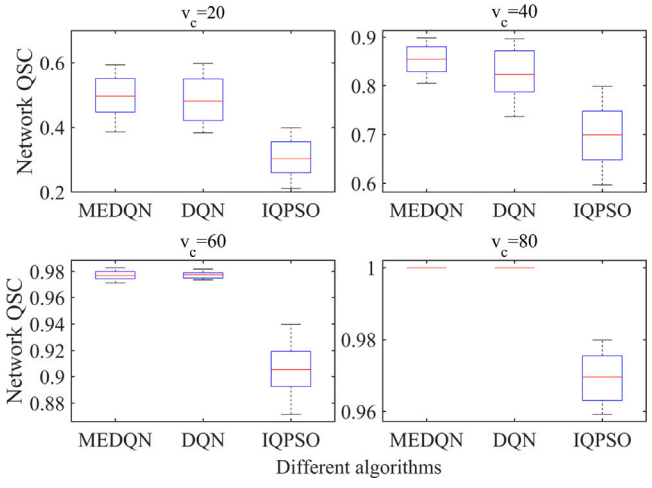

Fig. 7. Comparison of online and offline algorithms in terms of the network QSC by different MC charging powers.

Table 5
Comparison of different γ.

	0.91	0.92	0.93	0.94	0.95
QSC	0.898	0.904	0.902	0.885	0.907
TRs	5.43	3.252	7.70	4.940	8.40
SCR	0.903	0.912	0.907	0.889	0.912
NONs	5.921	6.388	5.806	6.578	**5.623**
	0.96	0.97	0.98	0.99	
QSC	0.906	**0.911**	0.903	0.887	
TRs	5.40	**9.18**	7.15	−1.37	
SCR	0.912	**0.916**	0.909	0.892	
NONs	5.929	5.643	5.899	6.207	

4.3.2. Return discount factor

The reward is the only feedback after the agent interacts with the environment and is the basis for the agent's learning. γ determines MC's emphasis on the return of future charging actions and directly affects the learning quality of the optimal mobile charging strategy. In this part, we compare the effects of different γ on the TRs of MEDQN and find the optimal γ for OMCS-QSC.

Table 5 shows that $\gamma = 0.97$ can obtain the largest TRs of about 9.18 and the optimal network QSC of about 0.911. From (4) and (5), γ decreases exponentially in the MC exploration progresses. MC only focuses on the real-time r_N until $\gamma = 0$. The goal of MC is to 'think ahead and globally' and learn the global optimal charging strategy. When $\gamma \leq 0.97$, the future return has little influence on the current action, MC cannot accurately estimate the $Q - evl$ function, thus decreasing TRs. Theoretically, NONs decreases, and QSC increases. When $\gamma = 0.95$, NONs is the lowest, as well as TRs and SCR. When $\gamma \geq 0.97$, MC becomes cautious about future charging steps, which may limit the learning vision of MEDQN.

4.3.3. Reward penalty factor

In the reward setting of MEDQN, as the node's remaining energy directly affects the node survival rate, the reward penalty factor β is vital in evaluating MC charging action. In this part, we compare the effects of different β on the performance of MEDQN and find the optimal β for OMCS-QSC. As shown in Table 6, when β increases from 1 to 2, TRs generally shows an upward trend, from about −11.24 to about 9.384. However, as β keeps increasing, TRs decreases, finally dropping

Table 6
Comparison of different β.

	1	1.2	1.4	1.6	1.8
QSC	0.861	0.876	0.862	0.871	0.874
TRs	−11.24	1.031	1.976	0.215	7.038
SCR	0.867	0.883	0.868	0.877	0.879
NONs	8.364	7.641	7.897	7.808	7.369
	2	2.2	2.4	2.6	
QSC	0.877	**0.884**	0.862	0.858	
TRs	**9.384**	8.469	8.204	−5.645	
SCR	0.881	**0.890**	0.864	0.860	
NONs	9.384	8.469	**8.204**	8.645	

to −5.645 at $\beta = 2.6$. It is worth noting that when $\beta = 2.2$, TRs is about 8.469 and not the largest, but the network QSC is optimal. When $\beta \leq 2$, there is a lesser punishment on e_r, and MC cannot charge the nodes with a larger contribution to the network, resulting in a lower network QSC.

In summary, we should reasonably choose different α, γ and β according to the needs of different problems, respectively, to reduce the fluctuation of the state–action value function in iterations, improve the agent's exploration efficiency and avoid the 'myopia' and 'hyperopia' of the agent.

5. Comparative simulation analysis

In this section, we will study the effects on network QSC of network scale, charging request energy threshold percentage, MC charging power, MC battery capacity and nodes' battery capacity, respectively. To evaluate the proposed MEDQN algorithm, we compare MEDQN with other four online algorithms, including the original DQN algorithm, and we run each experiment 20 times and take the average value as the final experimental result.

5.1. Simulation details

We assumed that 50 nodes are randomly and uniformly distributed in a fixed (100 m × 100 m) 2D monitoring region. Each node is powered by a 3.7 V/450 mAh alkaline rechargeable battery with the capacity e_m = 3.7 V×0.45 A×3600 s×24 h = 144 kJ. The initial residual energy e_{in} of each node is randomly generated between 43.2 kJ/s and 144 kJ/s, and the energy consumption rate v_{cs} of each node is randomly generated between 0.2 J/s and 1.8 J/s under each CTS. MC performs the charging task when one of the nodes in WRSN sends the charging request to MC. MC battery capacity E_m is limited and set at Ne_m = 72000. In addition, the MC moving speed v_m is set at 5 m/s, the MC charging power v_c is set at 40 J/s, and the energy consumption rate of the moving unit distance v_l is set at 10 J/m. More simulation parameters are shown in Table 3. We implement the learning algorithm in Matlab2021b software on the workstation with a quad-core 3.1 GHz CPU and four NVIDIA GeForce GTX 3080.

5.2. Comparison algorithms

This section compares the performance of MEDQN with the Nearest-Job-Next with Preemption (NJNP) [17] and Temporal–Spatial Charging scheduling Algorithm (TSCA) [20], where NJNP finds the nearest requested charging node in space as the next charging point and optimizes the charging efficiency by saving mobile energy consumption; TSCA firstly builds a mobile charging sequence for the nodes that have sent charging requests, then it deletes inefficient nodes based on the real-time mobile efficiency value ($a_1(k)/d(k)$) and finally adds nodes with more significant sensing coverage contributions considering the coverage efficiency value of real-time nodes ($a_1(k)/e_r(k)$). Because the sensing coverage contribution of each node will affect network QSC, we also designed a comparison Maximum Sensing Contribution Greedy (MSCG) algorithm. MSCG sorts the real-time $r_k = a_1(k)/a_m - \beta e_r(k)/e_m - a_l(k)/a_{tm}$ of all nodes with charging requests from large to small and continuously charges the node in the contribution sequence. To highlight the advantages of MEDQN, we also compared it with the original DQN algorithm.

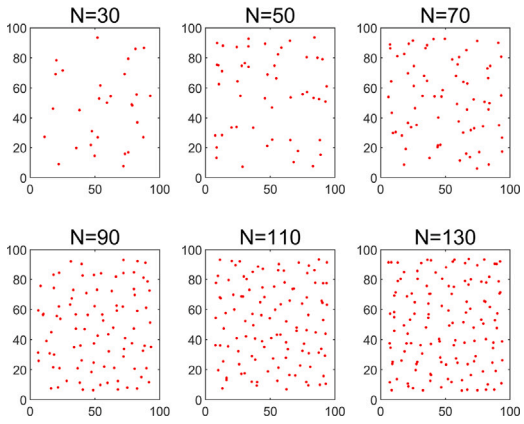

Fig. 8. Different network scales.

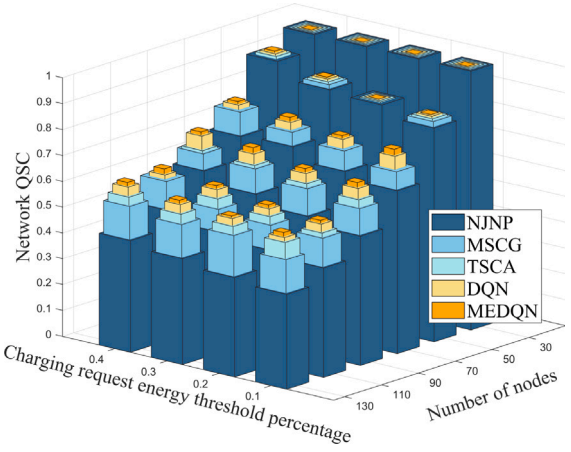

Fig. 9. Comparison of five algorithms in terms of the network QSC by deploying different numbers of nodes and charging request energy threshold percentages.

5.3. Comparison against different network scales

Within the same network monitoring region, increasing the number of nodes N directly affects the MC mobile charging decision and brings charging pressure to MC. In this section, we analyze the changes in network QSC for different N, charging request energy threshold percentages and MC charging powers. The specific random distributions of different N are shown in Fig. 8.

As shown in Figs. 9 and 10, as N increases under the same r_p, the network QSCs of all algorithms almost linearly decrease. As MC charging power is limited, the larger N, the more off-working nodes that cannot be charged in time due to energy depletion during MC charging, and the lower node survival ratio. In Fig. 9, as r_p increases under the same N, the initial remaining energy of all nodes is higher, and the overall working time of the network increases, MC has a high probability of charging the node with the lowest energy. Therefore, the network QSC of all algorithms improves.

As shown in Fig. 10, the network QSCs of five algorithms have the same increasing trend with the increase of v_c. Due to $t_c = (e_m - e_r)/v_r$, an increase of v_c reduces t_c, increase the number of charged nodes, improves the survival rate of nodes and sensing coverage ratio of

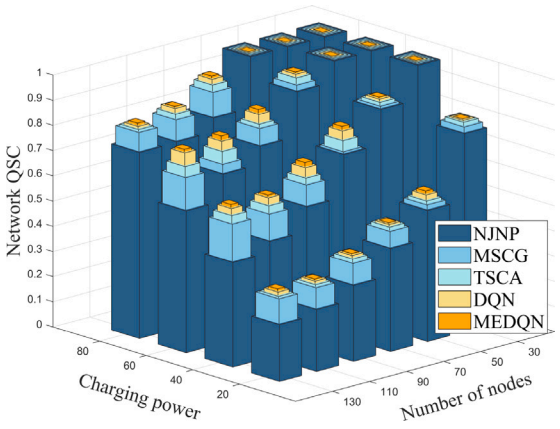

Fig. 10. Comparison of five algorithms in terms of the network QSC by deploying different numbers of nodes and charging powers.

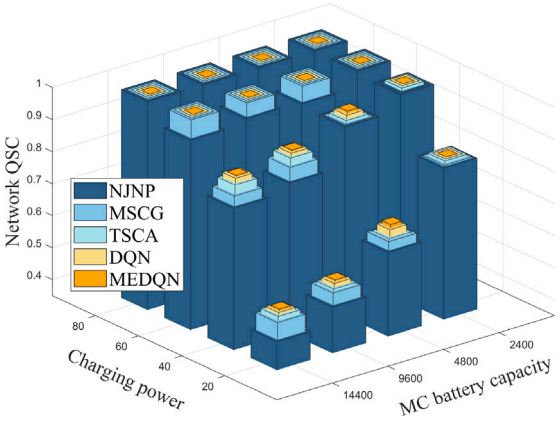

Fig. 11. Comparison of five algorithms in terms of the network QSC by deploying different MC carrying capacities and charging powers.

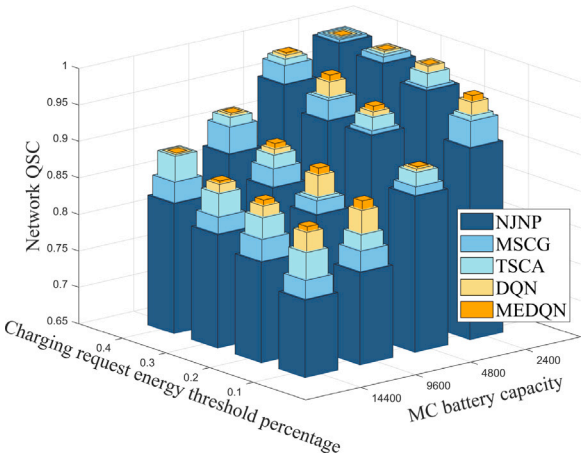

Fig. 12. Comparison of five algorithms in terms of network QSC by deploying different MC carrying capacities and charging request energy threshold percentages.

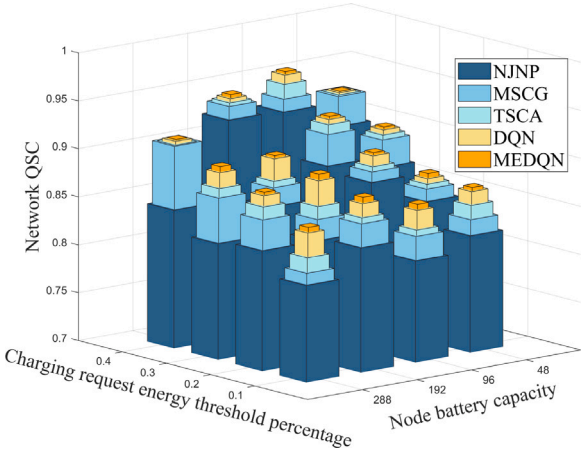

Fig. 13. Comparison of five algorithms in terms of network QSC by deploying different node battery capacities and charging request energy threshold percentages.

WRSN. When v_c is close to or greater than N, MC charging capability can almost meet the overall energy consumption rate of WRSN, and WRSN may achieve the optimal network QSC via a reasonable charging strategy. Summarizing the comparison above, MEDQN is superior to other algorithms in optimizing network QSC.

5.4. Comparison against different MC battery capacities

MC battery capacity affects the charging cycle and the number of charging nodes. In this part, we analyze the changes in network QSC of five algorithms for different MC battery capacities E_m, different MC charging powers and charging request energy threshold percentages.

As shown in Fig. 11, under the same v_c, increasing E_m can prolong the charging cycle but decrease the network QSC. Due to the limitation of MC charging capability, the nodes that have been charged will have insufficient energy again in the extended charging cycle, which increases NONs and reduces SCR. Therefore, we can conclude that if v_c is limited and cannot meet the whole network consumption rate, it is wise to appropriately reduce E_m to charge WRSN in multiple short charging cycles.

As shown in Fig. 12, under the same E_m, the network QSC increases as r_p increases until it reaches the optimum. The increase in r_p means that when MC performs the charging task, E_{in} is relatively high, and t_c can be effectively shortened. Because of $t_c \gg t_l$, when the MC charging capacity is limited, nodes may stop working in t_c with a higher probability, so NONs is reduced. When $E_m = 14400$ and $r_p = 0.1$, the network QSC reaches the minimum of 0.78. Because E_{in} of nodes are low, in the initial stage of MC performing the charging task, many nodes will stop working simultaneously. Therefore, in practical applications, we should set a larger r_p as much as possible to increase E_{in} to maximize the network QSC.

5.5. Comparison against different node battery capacities

The node battery capacity e_m will have a double-sided impact on the network sensing coverage performance because it affects both the working time of nodes and t_c. In this part, we analyze the changes in network QSC of four algorithms for different node battery capacities, different MC charging powers and charging request energy threshold percentages.

In Fig. 13, with the increase of e_m under the same r_p, the overall network QSC shows a fluctuating trend. The extended working time of nodes and t_c does not mean that the network QSC can be improved. When MC charging capability is limited, increasing e_m is equivalent to the reduction of v_c, indicating that the time needed for MC to charge the node increases, fewer nodes may not be charged in time, negatively impacting network QSC in the charging cycle. Therefore, in practical applications, we need to set the battery capacity of the node according to the actual MC charging capability and charging demands of WRSN.

As shown in Fig. 14, with the increase of v_c, the network QSC under different e_m all increase significantly. Four algorithms can achieve optimal network sensing coverage when $v_c = 80$. However, it fluctuates

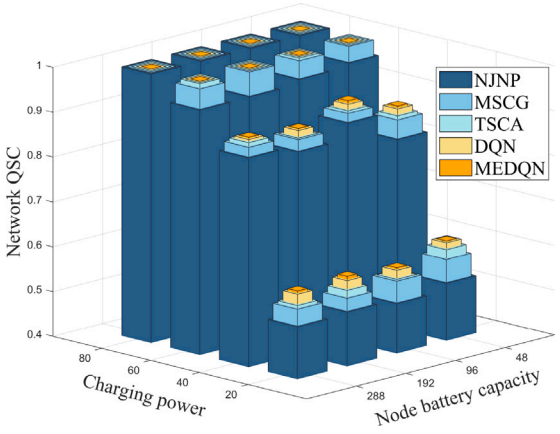

Fig. 14. Comparison of five algorithms in terms of network QSC by deploying different node battery capacities and charging powers.

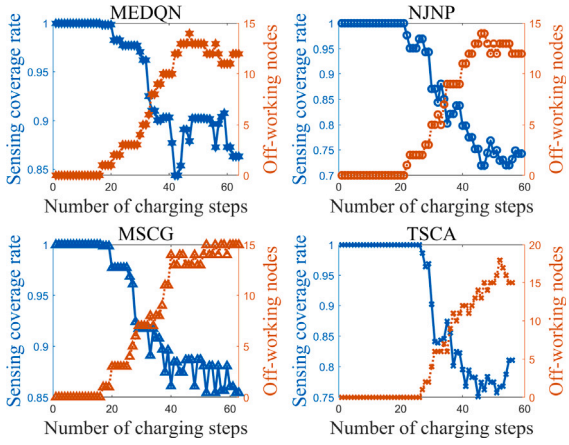

Fig. 15. Network SCR and NONs in the charging cycle of four algorithms.

wildly when the network QSC is between $v_c = 20$ and $v_c = 40$. Most nodes have no time to be charged during this period, and the network QSC is only about 0.55. However, there are 50 nodes in this experimental environment, and the average energy consumption of each node is 1 J/s. Therefore, $v_c = 40$ is close to the overall energy consumption of the network. Under a reasonable mobile charging scheduling strategy, the network QSC can reach a higher level.

5.6. Discussion the advantage of MEDQN

The comparison simulations performed in this study have essential scientific and practical significance, which can help us evaluate the effectiveness and applicability of the proposed MEDQN algorithm under different conditions. Finally, we analyze the reasons for the performance advantages of MEDQN.

Fig. 15 is the overall change of SCR and NONs in one charging cycle for the four algorithms. The comparison shows that before the 26th charging action, TSCA can guarantee the average network SCR to be 1. The reason is that TSCA always insists on charging nodes with lower e_r to reduce non before no nodes stop working. However, starting from the 27th charging action, the average sensing coverage of the network decreases almost linearly. Because MC charging capability is limited, MC is too late for all nodes whose remaining energy is about to be exhausted, and many nodes stop working simultaneously.

In contrast, MEDQN and MSCG consider the sensing coverage contribution of each node. Different from MSCG, which always charges the node with the most significant sensing coverage contribution, MEDQN can consider short-term and long-term benefits to optimize the network QSC in the charging cycle. MEDQN may not charge the node with the lowest e_r or one that has stopped, and it can choose the node with the lower e_r that has a more significant sensing coverage contribution to other nodes to ensure the overall network QSC in the charging cycle.

6. Conclusion

Considering the constraints of MC charging capability on the network QCS in dynamic WRSNs, this paper proposes a novel model-free MEDQN algorithm for OMCS-QSC to maximize the network QSC by finding the optimal charging strategy according to the real-time network state. To evaluate the MC real-time charging action, we design a novel reward function based on the real-time node's sensing coverage contribution. Extensive experiments are conducted to evaluate the performance of MEDQN for OMCS-QSC under different settings, including different network scales, MC charging powers, MC battery capacities, nodes' battery capacities, and the charging request energy threshold percentages. Experimental results show that MEDQN is superior to other online algorithms in maximizing network QSC, especially in large-scale WRSNs.

As the future work, based on the online approach proposed in this paper, designing an online algorithm for joint mobile charging scheduling and MC charging time control for the optimal network QSC in dynamic WRSNs should be considered.

CRediT authorship contribution statement

Jinglin Li: Writing – original draft, Software, Methodology, Investigation, Data curation, Conceptualization. **Haoran Wang:** Validation, Software, Investigation. **Chengpeng Jiang:** Validation, Data curation. **Wendong Xiao:** Validation, Supervision, Project administration, Conceptualization.

Declaration of competing 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.

Data availability

No data was used for the research described in the article.

Acknowledgments

This work was supported in part by the National Natural Science Foundations of China (NSFC) under Grant 62173032, the Foshan Science and Technology Innovation Special Project, China under Grant BK22BF005, and the Regional Joint Fund of the Guangdong Basic and Applied Basic Research Fund, China under Grant 2022A1515140109.

References

[1] Guangxu Zhu, Dongzhu Liu, Yuqing Du, Changsheng You, Jun Zhang, Kaibin Huang, Toward an intelligent edge: Wireless communication meets machine learning, IEEE Commun. Mag. 58 (1) (2020) 19–25.

[2] Youchao Wang, SM Shariar Morshed Rajib, Chris Collins, Bruce Grieve, Low-cost turbidity sensor for low-power wireless monitoring of fresh-water courses, IEEE Sens. J. 18 (11) (2018) 4689–4696.

[3] Yandong Zheng, Rongxing Lu, Songnian Zhang, Yunguo Guan, Jun Shao, Hui Zhu, Toward privacy-preserving healthcare monitoring based on time-series activities over cloud, IEEE Internet Things J. 9 (2) (2022) 1276–1288.

[4] Silvia Macis, Daniela Loi, Andrea Ulgheri, Danilo Pani, Giuliana Solinas, Serena La Manna, Vincenzo Cestone, Davide Guerri, Luigi Raffo, Design and usability assessment of a multi-device SOA-based telecare framework for the elderly, IEEE J. Biomed. Health Inf. 24 (1) (2020) 268–279.

[5] Fenghua Zhu, Yisheng Lv, Yuanyuan Chen, Xiao Wang, Gang Xiong, Fei-Yue Wang, Parallel transportation systems: Toward IoT-enabled smart urban traffic control and management, IEEE Trans. Intell. Transp. Syst. 21 (10) (2020) 4063–4071.

[6] Chao Zhang, Guanghui Zhou, Han Li, Yan Cao, Manufacturing blockchain of things for the configuration of a data- and knowledge-driven digital twin manufacturing cell, IEEE Internet Things J. 7 (12) (2020) 11884–11894.

[7] Riheng Jia, Xiuling Zhang, Yanju Feng, Tianliang Wang, Jianfeng Lu, Zhonglong Zheng, Minglu Li, Long-term energy collection in self-sustainable sensor networks: A deep Q-learning approach, IEEE Internet Things J. 8 (18) (2021) 14299–14307.

[8] Abhinav Tomar, Kumar Nitesh, Prasanta K. Jana, An efficient scheme for trajectory design of mobile chargers in wireless sensor networks, Wirel. Netw. 26 (2020) 897–912.

[9] Zhenchun Wei, Chengkai Xia, Xiaohui Yuan, Renhao Sun, Zengwei Lyu, Lei Shi, Jianjun Ji, The path planning scheme for joint charging and data collection in WRSNs: A multi-objective optimization method, J. Netw. Comput. Appl. 156 (2020) 102565.

[10] Chengpeng Jiang, Fen Liu, Jinglin Li, L.V. Peng, Wendong Xiao, Mobile energy replenishment scheduling based on quantum-behavior particle swarm optimization, in: 2020 39th Chinese Control Conference, CCC, IEEE, 2020, pp. 5253–5258.

[11] Madana Srinivas, Tarachand Amgoth, Delay-tolerant charging scheduling by multiple mobile chargers in wireless sensor network using hybrid GSFO, J. Ambient Intell. Humaniz. Comput. (2022) 1–17.

[12] Lei Mo, Angeliki Kritikakou, Shibo He, Energy-aware multiple mobile chargers coordination for wireless rechargeable sensor networks, IEEE Internet Things J. 6 (5) (2019) 8202–8214.

[13] Tang Liu, Baijun Wu, Shihao Zhang, Jian Peng, Wenzheng Xu, An effective multi-node charging scheme for wireless rechargeable sensor networks, in: IEEE INFOCOM 2020 - IEEE Conference on Computer Communications, 2020, pp. 2026–2035.

[14] Guangjie Han, Haofei Guan, Jiawei Wu, Sammy Chan, Lei Shu, Wenbo Zhang, An uneven cluster-based mobile charging algorithm for wireless rechargeable sensor networks, IEEE Syst. J. 13 (4) (2018) 3747–3758.

[15] Wenzheng Xu, Weifa Liang, Xiaohua Jia, Haibin Kan, Yinlong Xu, Xinming Zhang, Minimizing the maximum charging delay of multiple mobile chargers under the multi-node energy charging scheme, IEEE Trans. Mob. Comput. 20 (5) (2021) 1846–1861.

[16] Jinglin Li, Chengpeng Jiang, Jing Wang, Taian Xu, Wendong Xiao, Mobile charging sequence scheduling for optimal sensing coverage in wireless rechargeable sensor networks, Appl. Sci. 13 (5) (2023) 2840.

[17] Liang He, Linghe Kong, Yu Gu, Jianping Pan, Ting Zhu, Evaluating the on-demand mobile charging in wireless sensor networks, IEEE Trans. Mob. Comput. 14 (9) (2014) 1861–1875.

[18] Chi Lin, Yu Sun, Kai Wang, Zhunyue Chen, Bo Xu, Guowei Wu, Double warning thresholds for preemptive charging scheduling in wireless rechargeable sensor networks, Comput. Netw. 148 (2019) 72–87.

[19] Chi Lin, Ding Han, Jing Deng, Guowei Wu, p^2S:A primary and passer-by scheduling algorithm for on-demand charging architecture in wireless rechargeable sensor networks, IEEE Trans. Veh. Technol. 66 (9) (2017) 8047–8058.

[20] Chi Lin, Jingzhe Zhou, Chunyang Guo, Houbing Song, Guowei Wu, Mohammad S. Obaidat, TSCA: A temporal-spatial real-time charging scheduling algorithm for on-demand architecture in wireless rechargeable sensor networks, IEEE Trans. Mob. Comput. 17 (1) (2017) 211–224.

[21] Zhansheng Chen, Hong Shen, Xiaofan Zhao, Delay-tolerant on-demand mobile charging scheduling scheme for wireless rechargeable sensor networks, in: 2018 9th International Symposium on Parallel Architectures, Algorithms and Programming, PAAP, IEEE, 2018, pp. 29–35.

[22] Naween Kumar, Dinesh Dash, Mukesh Kumar, An efficient on-demand charging schedule method in rechargeable sensor networks, J. Ambient Intell. Humaniz. Comput. 12 (7) (2021) 8041–8058.

[23] Hongli Yu, Chih-Yung Chang, Yajun Wang, Diptendu Sinha Roy, Xing Bai, CAERM: Coverage aware energy replenishment mechanism using mobile charger in wireless sensor networks, IEEE Sens. J. 21 (20) (2021) 23682–23697.

[24] Yuanping Kan, Chih-Yung Chang, Chin-Hwa Kuo, Diptendu Sinha Roy, Coverage and connectivity aware energy charging mechanism using mobile charger for WRSNs, IEEE Syst. J. 16 (3) (2021) 3993–4004.

[25] Lintong Jiang, Xiaobing Wu, Guihai Chen, Yuling Li, Effective on-demand mobile charger scheduling for maximizing coverage in wireless rechargeable sensor networks, Mob. Netw. Appl. 19 (4) (2014) 543–551.

[26] Xiao Lu, Ping Wang, Dusit Niyato, Dong In Kim, Zhu Han, Wireless charging technologies: Fundamentals, standards, and network applications, IEEE Commun. Surv. Tutor. 18 (2) (2015) 1413–1452.

[27] Zhenchun Wei, Fei Liu, Zengwei Lyu, Xu Ding, Lei Shi, Chengkai Xia, Reinforcement learning for a novel mobile charging strategy in wireless rechargeable sensor networks, in: Wireless Algorithms, Systems, and Applications: 13th International Conference, WASA 2018, Tianjin, China, June 20-22, 2018, Proceedings 13, Springer, 2018, pp. 485–496.

[28] David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, et al., Mastering the game of go without human knowledge, Nature 550 (7676) (2017) 354–359.

[29] Phi Le Nguyen, Thanh-Hung Nguyen, Kien Nguyen, et al., Q-learning-based, optimized on-demand charging algorithm in WRSN, in: 2020 IEEE 19th International Symposium on Network Computing and Applications, NCA, IEEE, 2020, pp. 1–8.

[30] Xianbo Cao, Wenzheng Xu, Xuxun Liu, Jian Peng, Tang Liu, A deep reinforcement learning-based on-demand charging algorithm for wireless rechargeable sensor networks, Ad Hoc Netw. 110 (2021) 102278.

[31] Chengpeng Jiang, Ziyang Wang, Shuai Chen, Jinglin Li, Haoran Wang, Jinwei Xiang, Wendong Xiao, Attention-shared multi-agent actor–critic-based deep reinforcement learning approach for mobile charging dynamic scheduling in wireless rechargeable sensor networks, Entropy 24 (7) (2022) 965.

[32] Meiyi Yang, Nianbo Liu, Lin Zuo, Yong Feng, Minghui Liu, Haigang Gong, Ming Liu, Dynamic charging scheme problem with actor-critic reinforcement learning, IEEE Internet Things J. 8 (1) (2020) 370–380.

[33] Shuai Chen, Chengpeng Jiang, Jinglin Li, Jinwei Xiang, Wendong Xiao, Improved deep Q-network for user-side battery energy storage charging and discharging strategy in industrial parks, Entropy 23 (10) (2021) 1311.

[34] Bhargavi Dande, Shi-Yong Chen, Huan-Chao Keh, Shin-Jer Yang, Diptendu Sinha Roy, Coverage-aware recharging scheduling using mobile charger in wireless sensor networks, IEEE Access 9 (2021) 87318–87331.

[35] André L.C. Ottoni, Erivelton G. Nepomuceno, Marcos S. de Oliveira, Daniela C.R. de Oliveira, Reinforcement learning for the traveling salesman problem with refueling, Complex Intell. Syst. 8 (3) (2022) 2001–2015.

[36] Santosh Soni, Manish Shrivastava, Novel wireless charging algorithms to charge mobile wireless sensor network by using reinforcement learning, SN Appl. Sci. 1 (2019) 1–18.

[37] Wendong Xiao, Lihua Xie, Jianyong Lin, Jianing Li, Multi-sensor scheduling for reliable target tracking in wireless sensor networks, in: 2006 6th International Conference on ITS Telecommunications, IEEE, 2006, pp. 996–1000.

[38] Qinglai Wei, Feiyue Wang, Reinforcement Learning, Tsinghua University Press, Beijing, 2022.

Jinglin Li received the B.E. degree in School of New Energy from Liaoning University of Technology, Jinzhou, China, in 2016 and received the M.E. degree in College of Information and Electrical Engineering from Shenyang Agricultural University, Shenyang, China, in 2018.

He is currently a Ph.D. student in the School of Automation and Electrical Engineering, University of Science and Technology Beijing, China. His research interests include reinforcement learning, mobile charging and wireless rechargeable sensor network.

Haoran Wang received the B.S. and M.S. degrees from the College of Electrical and Electronic Engineering, Henan Normal University, Xinxiang, China, in 2017 and 2021, respectively.

He is currently a Ph.D. student in the School of Automation and Electrical Engineering, University of Science and Technology Beijing, China. His research interests include reinforcement learning, wireless sensor networks.

Chengpeng Jiang received the B.S. degree in School of the Internet of Things from Harbin University of Science and Technology, Harbin, China, in 2015.

He is currently a Ph.D. student in the School of Automation and Electrical Engineering, University of Science and Technology Beijing, China. His research interests include reinforcement learning, mobile sensor scheduling, mobile charging and wireless rechargeable sensor network.

Wendong Xiao received the B.S. and Ph.D. degrees from Northeastern University, Shenyang, China, in 1990 and 1995, respectively. He held various academic and research positions with Northeastern University, POSCO Technical Research Laboratories, South Korea, Nanyang Technological University, Singapore, and the Institute for InfoComm Research, Agency for Science, Technology and Research (A* STAR), Singapore.

He is currently a Professor with the School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing, China. His research interests include wireless intelligent sensing, big data processing, energy harvesting based resource management, wearable computing for healthcare, wireless sensor networks, and Internet of Things.