Adaptive Monitoring Optimization Based on Deep-Q-Network for Energy Harvesting Wireless Sensor Networks

Xuecai Bao[1,2(✉)], Peilun Bian[1,2], Wenqun Tan[1,2], Xiaohua Xu[2], and Jugen Nie[1,2]

[1] Jiangxi Province Key Laboratory of Water Information Cooperative Sensing and Intelligent Processing, Nanchang Institute of Technology, Nanchang, Jiangxi 330099, China
Lx97821@nit.edu.cn

[2] Jiangxi Provincial Technology Innovation Center for Ecological Water Engineering in Poyang Lake Basin, Jiangxi 330029, China

Abstract. In order to improve the energy efficiency of environmental monitoring for energy harvesting wireless sensor networks (EH-WSNs) in remote areas and achieve energy-neutral operation, an adaptive monitoring and energy management optimization method of EH-WSNs based on deep Q network (DQN) algorithm is proposed. In this paper, aiming at EH-WSNs with single-hop cluster structure, we first present a more realistic energy model established by combining different climate characteristics. Then, the optimization model of maximizing long-term monitoring utility is formulated based on harvested energy constraints. We use deep Q network (DQN) to learn random and dynamic solar energy harvesting process on solar-powered sensor nodes and optimize the monitored performance of EH-WSNs through the replay memory mechanism and freezing parameter mechanism. Finally, we present an adaptive monitoring optimization method based DQN to achieve the long-term utility. Through simulation verification and comparative analysis, in different rainy weather environments, the proposed optimization algorithm has greatly improved in terms of average monitoring reward, monitoring interruption rate and energy overflow rate. Moreover, it also indicates that the proposed algorithm has effective adaptation to the random and dynamic solar energy arrival.

Keywords: Energy Harvesting WSNs · Adaptive Monitoring · Deep Q Network · Long-term Utility

1 Introduction

Monitoring technology based on wireless sensor network (WSN) is one of the effective solutions to help supervise pollution emissions and promote the management of ecological environment [1, 2]. However, continuous monitoring often requires huge energy consumption and causes network congestion, such as high frequency monitoring or image monitoring. Although some congestion control and packet reordering algorithms provide the solution for network congestion [3, 4], traditional sensor network monitoring

Y. Cao and X. Shao (Eds.): MONAMI 2022, LNICST 474, pp. 316–330, 2023.
https://doi.org/10.1007/978-3-031-32443-7_23

system based on limited battery power still causes interruption of monitoring, especially in remote areas [1], where frequent battery replacement is too expensive and impractical. How to reduce the interruption of monitoring and improve the monitoring utility in remote areas is one of the important problems to be solved at present.

WSNs can use different types of energy sources, such as solar energy [5], wind energy [6] and so on. The energy from these external environments can be converted into electrical energy for the monitoring node by different energy conversion devices. Consequently, WSNs based on energy harvesting provide a solution for solving the energy management problem of monitoring in remote areas to a certain extent [7, 8].

In recent years, some traditional solutions based on energy harvesting technique have been applied to address the problem of sensor node lifetime [9–11]. Considering the large amount of energy consumption during cluster head selection stage and unequal harvested energy among nodes in EH-WSNs, Ren and Yao proposed an energy-efficient cluster head selection scheme [12]. Based on some traditional routing protocols such as LEACH, the scheme classified nodes with different functions and effectively scheduled them to deal with the energy management problem. Xiong focused on how to increase the network lifetime while satisfying the full target coverage in a novel hybrid EH-WSN, and then proposed a two-phase lifetime-enhancing method to meet these requirements [13]. To reduce transmission delay and improve network throughput, Bengheni deployed an enhanced energy management scheme in EH-WSNs to improve the overall performance of network [14], and it introduced an energy threshold policy to ensure a balance between the energy consumption and energy harvesting ability for each sensor node. Besides, Qiu started from the transmission strategy management in EH-WSNs, and used Lyapunov optimization theory to maximize the expected bits per packet transmission for source node in system [15].

However, solar energy has instability and random dynamic characteristics and cannot be controlled [16]. It is one of the major challenges for its energy management in EH-WSNs. Specially, it is very important to seek effective optimization strategies and realize efficient energy management of sensor networks for improving continuous environmental monitoring and prolonging network lifetime. Although many energy management methods were proposed to improve the network performance, most of them assumed that harvested energy was known in advance [17–20].Therefore, in order to adapt the random and dynamic of solar energy, we propose a novel DQN-based adaptive monitoring optimization method for energy management in EH-WSNs. In the proposed method, we first present the dynamic energy model, consumed energy model and the network model of cluster structure. Then, we formulate the problem of adaptive monitoring in EH-WSNs. Considering the dynamic characteristic of energy and excessive energy state, the DQN algorithm is utilize to solve the problem and improve the utility of the whole network.

The rest of this paper is organized as follows. Section 2 and Sect. 3 present the model assumptions and optimization problem formulation, respectively. The detailed adaptive monitoring optimization for EH-WSNs based on solar energy harvesting is presented in Sect. 4. The simulation verification and analysis are presented in Sect. 5. Finally, conclusions are derived in Sect. 6.

2 System Model

In this section, we first present the EH-WSN model based on cluster structure, then the energy consumption model and energy harvesting model are described.

2.1 Network Model

At present, the topology of WSNs is generally divided into two types: single-hop cluster structure and multi-hop Mesh network. For environment image monitoring in remote areas, due to the relatively high bandwidth requirement, the cluster network topology with a lower delay is more suitable than the mesh network. Hence, we describe our network model as shown in Fig. 1.

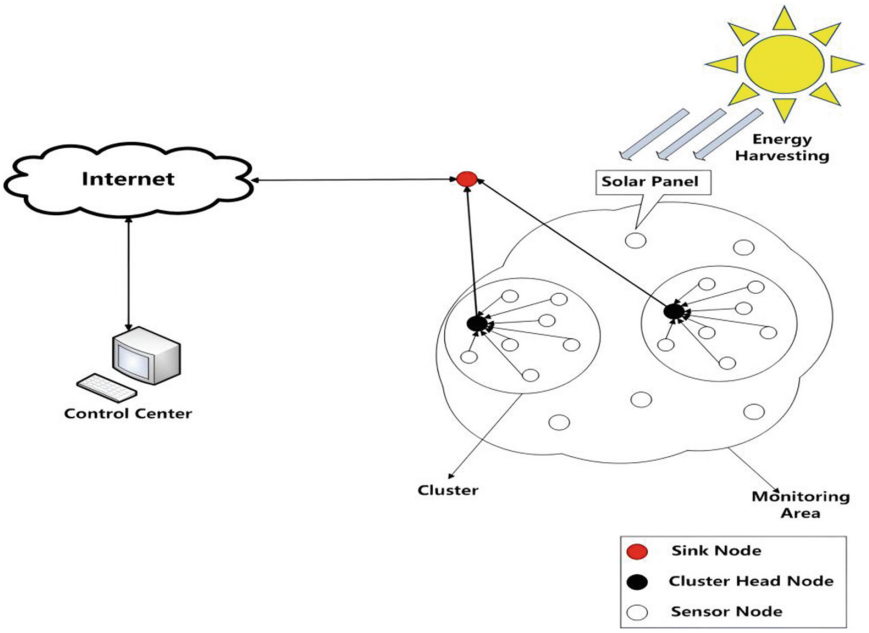


Fig. 1. Network model

2.2 Energy Consumption Model

In WSNs, the energy consumption of nodes is mainly composed of environmental monitoring, data transmission and data reception. Combined with two models of energy consumption [16, 20], the energy consumption models are defined as follows

$$E_{cons} = E_M + E_T + E_R \tag{1}$$

$$E_M = E_m * M\,\mathrm{s} * T \tag{2}$$

$$E_T = E_{elec} * l + \xi_{amp} * l * d^2 \tag{3}$$

$$E_R = E_{elec} * l \tag{4}$$

where E_{cons} is the total consumption energy, E_M is the consumption energy of monitoring, E_T and E_R are the energy consumed by transmitting and receiving 1bit data between two nodes with a distance of d, respectively. In (2), E_m is the energy consumed by once monitoring, which is a fixed value and M_s is the monitoring frequency of each time slot. Moreover, T refers to the total time slots of monitoring. It can be seen that the total energy consumption of environmental monitoring is basically proportional to the monitoring frequency. With the increase of monitoring frequency, the energy consumption of sensor nodes will also increase. In (3)–(4), E_{elec} denotes the energy dissipated per bit. d refers to the distance between sending node and receiving node and ξ_{amp} is the energy consumed by amplifier, which depends on the specification of sending amplifier. In this paper, we focus on the adaptive monitoring frequency optimization of sensor nodes and improve the total energy efficiency of monitoring.

2.3 Energy Harvesting Model

In this paper, we consider the solar energy as the harvested energy. For solar energy, solar panels are used to convert solar radiation into electrical energy to power the rechargeable battery in a node, and the battery provides energy for sensor node through energy management chip. Although solar energy is difficult to control, it is not completely unpredictable. Therefore, according to the model used in Lee and Zairi [16, 21], our model of solar energy harvesting process is defined as

$$E_H = \min\left\{E_{bc}, \eta \int_{\tau}^{\tau+h} p(t)dt\right\} \tag{5}$$

where E_H is the total harvested energy, E_{bc} is the maximum capacity of battery in sensor node, η is the uncertainty factors affecting solar energy harvesting, such as climate and weather in the environment. τ and $\tau+h$ represent the duration of solar energy harvesting in a day, where h is the execution time of energy harvesting, and the energy harvesting can be considered to obey Poisson distribution according to Lee [15]. *p(t)* is the probability density function of random process of solar energy harvesting in this period of time.

3 Problem Formulation

To improve the continuous monitoring and long-term survival of EH-WSNs in remote areas, we formulate the adaptive monitoring optimization based on the above system model to achieve trade-off between energy consumption and monitoring frequency, and the optimization problem can be described as maximizing the long-term utility

of network through the adjustment of monitoring frequency for each time slot. The optimization problem is written as follows.

$$\max(\lim_{T\to\infty}\sum_{t=0}^{T} r_t(M_s)) \tag{6}$$

$$s.t.\ E_{res}(t+1) = \min\{E_H(t) + E_{res}(t) - E_{cons}(t), E_{bc}\} \tag{7}$$

$$E_H(t) \geqslant 0 \tag{8}$$

$$0 \leqslant E_{res}(t) \leqslant E_{bc} \tag{9}$$

$$0 \leqslant E_{cons}(t) \leqslant E_{res}(t) \tag{10}$$

The (6) denotes the optimization objection is to maximize cumulative environmental monitoring reward obtained over a period of time slots T, where $r_t(M_s)$ is the instant reward obtained through the optimized monitoring frequency M_s of sensor node in each time slot t under the premise of available residual energy. The (7)–(10) are the constrained conditions, where $E_H(t)$ is the harvested energy of nodes, $E_{res}(t)$ is the residual energy of nodes at current time slot t, $E_{cons}(t)$ is the energy consumed by nodes at time slot t, and E_{bc} is the total battery capacity of sensor nodes. In (7), the final value for the sum of $E_H(t)$ and $E_{res}(t)$ minus $E_{cons}(t)$ cannot exceed total recharge battery capacity of nodes.

Aiming at the optimization problem, traditional optimization methods are difficult to solve this complex optimization problem. The DQN algorithm, which combines the advantages of reinforcement learning and deep learning, can optimize the long-term utility and take multi-state dimension into consideration.

4 Adaptive Monitoring Optimization Method Based DQN

4.1 Algorithm Principle

In Q-learning algorithm, the Q value table is usually used to store the Q value obtained by taking different action a under each state s. However, this approach usually encounters dimension disaster problems in dealing with large or even continuous tasks. In order to better solve this problem, a function Q_N composed of parameters ω is introduced to approximate the Q value, namely value function approximation, as shown below.

$$Q_N(\boldsymbol{s}, \boldsymbol{a}; \omega) \approx Q(s, a) \tag{11}$$

where $\boldsymbol{s}$ and $\boldsymbol{a}$ are vector representations of state s and action a, respectively. And with the development of deep learning, neural network technology and the value function approximation show good compatibility. Therefore, the DQN algorithm comes into being through the combination of deep learning and Q-learning. Compared with Q-learning, the obvious feature of DQN algorithm is to convert Q-function into a neural network through the value function approximation.

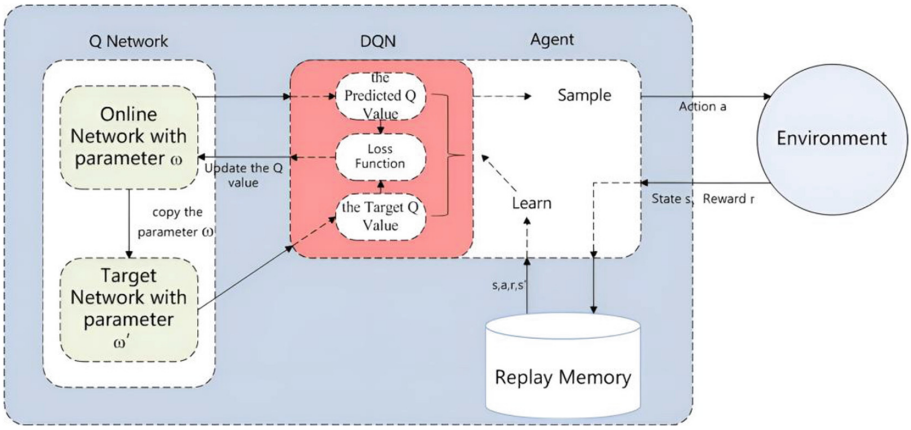


Fig. 2. DQN algorithm training process with replay memory and target network

Compared with Q-learning algorithm, DQN algorithm also has two new mechanisms, namely, replay memory and target network, as shown in Fig. 2.

To be specific, the first mechanism is to build replay memory to store the data obtained by interacting with the environment. When DQN algorithm is updated, it only needs to extract the previous experience data for learning. The second mechanism is to use the target network to freeze parameters ω. Specifically, the parameters ω are copied from the online network to target network at regular intervals by setting the appropriate update frequency. And in process of network training, only online network needs to be updated in real time. Therefore, when only the parameters of online network need to be adjusted, it becomes a regression problem. For two networks, the closer the target Q value is to the predicted Q value, the better the result is, so it is necessary to minimize its mean square error, which is defined as follows

$$L = (y - Q(\boldsymbol{s}, \boldsymbol{a}; \omega))^2 \tag{12}$$

where y and $Q(\boldsymbol{s}, \boldsymbol{a}; \omega)$ represent target Q value and predicted Q value respectively. Predicted Q value needs to be continuously updated by online network. The update process is similar to the Q-function update in Q-learning algorithm, and Q-function is updated as

$$Q_{t+1}(s_t, a_t) = Q_t(s_t, a_t) + \alpha\left[r_{t+1} + \gamma \max Q(s_{t+1}, a') - Q_t(s_t, a_t)\right] \tag{13}$$

Then parameters in target network are copied from online network, and the target Q value in target network can be obtain as

$$y = \begin{cases} r, if\_end\ is\ true \\ r + \gamma \max Q'(s', a'; \omega'), if\_end\ is\ false \end{cases} \tag{14}$$

where *if_end* is a sign to judge whether the algorithm ends.

4.2 State and Action Space

According to the network model in Sect. 2, sensor nodes generally need to send some state information to the sink node(it refers to agent in Fig. 4), such as residual energy, harvested energy and current time slot for a day or night. Therefore, the final state space $\boldsymbol{s}$ at time slot t is defined as

$$\boldsymbol{s} = [E_{res}(t), E_H(t), N(t)] \tag{15}$$

where $E_{res}(t)$ is the residual energy of nodes at time slot t, $E_H(t)$ is the energy harvested by nodes through external environment at time slot t, $N(t)$ is to judge whether the current time slot t is in the daytime or night.

After agent obtains the state of environment, it will select an action from the given action space $\boldsymbol{a}$ according to the strategy. This paper focuses on optimizing the action strategy of monitoring frequency to improve energy efficiency and monitoring utility of EH-WSNs. Therefore, we set monitoring frequency $M_s(t)$ in each time slot as the action. Since continuous action cannot be processed, the action space needs to be discretized to reduce the convergence time of the algorithm. Therefore, the discretized action space can be written as follows.

$$\boldsymbol{a}_D = [M_s(t)|d\,] \tag{16}$$

where d is the interval of discretization. The larger the interval of the discretization, the fewer actions contained in the action space. On the contrary, the smaller the interval, the more actions the action space can describe. Accordingly, assume that there is an action space set with A actions, namely $\boldsymbol{a}_D = \{0, 1, 2, \ldots, n, \ldots, A-1\}$, 0 means that nodes enter sleep, and n means that nodes monitor n times in each time slot.

4.3 Reward Function

According to the description of the optimization model, the setting of reward function needs to consider two requirements: the first is to utilize the harvested energy of sensor nodes to improve long-term utility by increasing the number of monitoring times in each time slot as much as possible; the second is to avoid the situation that causes monitoring interruption of sensor nodes in many time slots due to insufficient residual energy.

Then, according to the energy model proposed in Sect. 2 and three-stage energy management strategy proposed in [6], we consider the differences of the energy harvested at different time period in different environments. Accordingly, we design the different reward functions for different state to optimize the action selection of nodes through the feedback of agent. In the designed reward function, harvested energy and different environments are added to help nodes to make better decisions. Moreover, we also make corresponding distinctions for the residual energy states in different intervals based on sigmoid curve and Mexican hat curve. Finally, we define reward functions r_c and r_s based on the different weather state W. The r_c and r_s represent the instant rewards obtained by agent in rainy and sunny days, respectively. The specific expressions are defined as follows

$$r_c = \begin{cases} c((1-s_3)(\frac{4}{(1+\exp(-ba))(1+\exp(ba))} - 1) + s_3(\frac{2}{1+\exp(-b\frac{s_1}{E_{bc}})} - 1)), a \neq 0, s_1 \in (0, \frac{E_{bc}}{6}] \\ c((1-s_3)(\frac{4}{(1+\exp(-ba))(1+\exp(ba))} - 1) + s_3(\frac{2}{1+\exp(-bs_2)} - 1)), a \neq 0, s_1 \in (\frac{E_{bc}}{6}, \frac{E_{bc}}{2}] \\ c((1-s_3)(\frac{4}{(1+\exp(-ba))(1+\exp(ba))} - 1) + s_3(\frac{2}{1+\exp(-b(s_2+\frac{s_1}{E_{bc}}))} - 1)), a \neq 0, s_1 \in (\frac{E_{bc}}{2}, E_{bc}] \\ -r_{\max}, a \neq 0, s_1 = 0 \\ 0, a = 0 \end{cases} \tag{17}$$

$$r_s = \begin{cases} c((1-s_3)(\frac{4}{(1+\exp(-ba))(1+\exp(ba))} - 1) + s_2(\frac{2}{1+\exp(-bs_2)} - 1)), a \neq 0, s_1 \in (0, \frac{E_{bc}}{6}] \\ c((1-s_3)(\frac{4}{(1+\exp(-ba))(1+\exp(ba))} - 1) + s_2(\frac{2}{1+\exp(-b(s_1+a))} - 1)), a \neq 0, s_1 \in (\frac{E_{bc}}{6}, \frac{E_{bc}}{2}] \\ -r_{\max}, a \neq 0, s_1 = 0 \\ 0, a = 0 \end{cases} \tag{18}$$

where a is the action, E_{bc} is the total battery capacity of nodes, s_1, s_2 and s_3 in (17) and (18) represent the first state $E_{res}(t)$, the second state $E_H(t)$ and the third state $N(t)$ in current time slot t, respectively. Moreover, $s_3 = 1$ means daytime and $s_3 = 0$ means night. Therefore, in the daytime, only the second half of formula is calculated, i.e., the sigmoid-like function. In the night, only the first half is calculated, i.e., the Mexican hat-like curve function. Furthermore, c and b represent the control of the amplitude and slope of the function, respectively. According to [6], c is set to 2 and b is set to 1 in (17) and (18). In rainy days, agent obtains different rewards based on the ratio of residual energy of nodes to total battery capacity and harvested energy. In sunny days, the rewards obtained by agent depend on harvested energy and action. At the same time, during the night, agent only obtains the reward value according to action. The larger the action taken, the smaller the reward received by agent. And the agent can avoid the excessive energy consumption of nodes through this negative feedback.

In addition, in order to further meet the energy storage constraints of nodes and avoid large-scale energy depletion of nodes, a penalty term $-r_{max}$ can help nodes constrain action decisions in different environments, where r_{max} is the maximum instant reward from current environment. This setting is to achieve that the instant rewards is imposed a heavy penalty when energy depletion of nodes happens. The specific description of the proposed DQN-based adaptive monitoring optimization algorithm for EH-WSNs is presented in **Algorithm 1**.

Next, the detail description of **Algorithm 1** is as follows. The initialization of a series of parameters and the corresponding setting is implemented in line 1–6. In line 7, the controller of sensor node sends the state to the agent, i.e., the deep Q network, and then the deep Q network feeds back the random or optimal action for current time slot t according to the ε-greedy policy, which is defined as

$$a_t = \begin{cases} \arg\max Q(s_t, a_t, \omega), \textit{ if } p \geqslant \varepsilon \\ \textit{RandomAction}, \quad \textit{if } p < \varepsilon \end{cases} \tag{19}$$

where *RandonAction* represents randomly selecting an action from the action space, ε is the greedy degree we set. The ε-greedy policy is utilized to balance the exploration and

exploitation, i.e., to balance the reward maximization based on the knowledge already known with trying new actions to obtain unknown knowledge.

After obtaining the action, the system will transfer to the next state when an action is performed (line8), where the residual energy of nodes $E_{res}(t)$ will change with the energy consumed by the action selection of the current time slot, the second state $E_H(t)$ will change regularly according to the energy harvesting model proposed in Sect. 2.3 and the last state $N(t)$ also depends on the change of time.

Algorithm1: DQN-based Adaptive Monitoring Optimization Algorithm

1. **Initialization:** Initialize the online network Q with parameters ω, the target network Q' with parameters ω', the final episode M and the replay memory D
2. **while** episode $k \leq M$ **do**
3. Initialize the beginning state space $\boldsymbol{s}$ and current weather W.
4. **For** $t = 1, 2 \ldots, T+1$ **do**
5. Set environment as Env according to different weather characteristics.
6. Choose a random probability p.
7. Choose action a_t by using ε-greedy policy and execute it.
8. Obtain the next state s_{t+1} based on the selected action and current environment.
9. Calculate reward according to formula (18) and (19) , i.e. $r_t = \begin{cases} r_c \leftarrow Env(a_t), if\ W = 0 \\ r_s \leftarrow Env(a_t), if\ W = 1 \end{cases}$.
10. Store the experience $\{s_t, a_t, r_t, s_{t+1}\}$ in replay memory D.
11. Get a batch of samples $\{s_t, a_t, r_t, s_{t+1}\}$ from the replay memory D.
12. Calculate $y = \begin{cases} r_t, if\ t \text{ terminates } at\ step\ T+1 \\ r_t + \gamma \max Q'(s_{t+1}, a'_t; \omega'), otherwise \end{cases}$.
13. Perform a gradient descent step by using $(y - Q(s_t, a_t; \omega))^2$ to update the parameters ω of online network.
14. Return the value of parameters ω in the online network Q.
15. Every N times update target network parameters ω'.
16. episode $k = k+1$
17. **End For**
18. **End while**
19. Output the target network Q'

Then the reward can be calculated according to the reward function (line 9). And it should be noted that when agent interacts with the environment, if the current weather is rainy, i.e., $W = 0$, the instant reward r_c can be obtained according to formula (17) above, and if the current weather is sunny, i.e., $W = 1$, the instant reward r_s can be obtained according to formula (18), both rewards depend on residual energy and harvested energy.

Inside the network, the replay memory stores agent's experience of each time slot (line 10). The parameters ω of online network are updated every time with samples from

the replay memory (line 11–14), and finally the parameters ω' of target network are copied from the online network every N times (line 15).

5 Performance Evaluation

5.1 Simulation Settings

In our simulations, we used a GPU-based server with the software environment of TensorFlow 1.3.0 and Python 3.6. Thereafter, we investigate the performance of our proposed DQN-based adaptive monitoring optimization algorithm (Proposed DQN) compared with the Q-learning algorithm, random method and greedy algorithm [22] (Greedy).

Here, it should be noted that the rainy environment means that the number of rainy days is not less than that of sunny days during a period time. Simultaneously, we compare the performance of proposed algorithm with the other three methods and investigate the performance metrics related to average monitoring reward, monitoring interruption rate and energy overflow rate based on the same energy harvesting model. The detail parameter settings are summarized in Table 1.

Table 1. Parameter values used in the simulations

Parameter	Value	Description
Discount factor γ	0.9	Discount factor used for the DQN algorithm
Exploration ε	0.1	Chance of random action in the ε-greedy exploration
Learning rate α	0.0001	Learning rate used by AdamOptimizer
Total episode k	5000	How many episodes are used to train the network model
Replay memory size	1000	Size of the container for storing learning experiences
Target network update frequency N	10	Rate to update target Q network towards online Q network
Batch size	32	How many learning experience are used for each episode
Battery capacity E_{bc}	120	Maximum capacity of rechargeable battery for sensor node
Energy consumption units E_{con}	0 to 4	Range of energy consumption units per time slot by monitoring
Harvested energy in sunny days E_H	0 to 9	Random range of harvested energy units per time slot in sunny days
Harvested energy in rainy days E_H	0 to 5	Random Range of harvested energy units per time slot in rainy days

In Table 1, we list the parameter values of harvested energy E_H in two different weathers and energy consumption units E_{cons} in different monitoring frequency. For action space $\boldsymbol{a}$, we also set four actions, i.e., the node monitors 0 to 4 times in each time slot and consumes the corresponding amount of energy. Subsequently, since there is a certain gap for energy harvesting in different environments, we set two different ranges of values. In sunny days, due to the high solar radiation intensity, the range of harvested energy is set to 1 to 9 energy units; on the contrary, for the rainy days, the range is set to 1 to 5 energy units. If there is no light at night, the value of harvested energy is 0.

5.2 Results and Comparative Analysis

For environmental monitoring of EH-WSN in remote areas, the performance of the rainy environment is the most concerned about the optimization problem. Therefore, according to the simulation scenarios in above subsection, we center on the performance analysis of the proposed algorithm compared to other three algorithms under the rainy environment with two different ratio of rainy days to sunny days. Here, the rainy environment can better reflect the trade-off performance between monitoring frequency and energy consumption.

To verify the performance of proposed DQN-based algorithm, we present the results for rainy days accounted for 70%. The corresponding state-action result is shown in Fig. 3

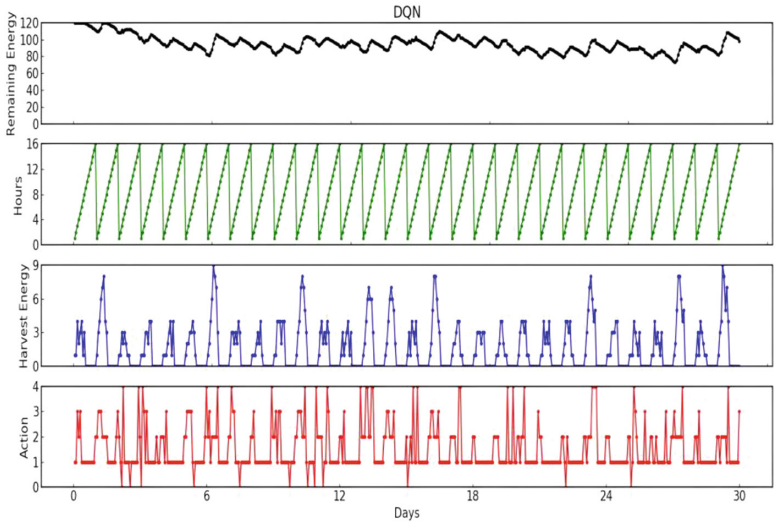
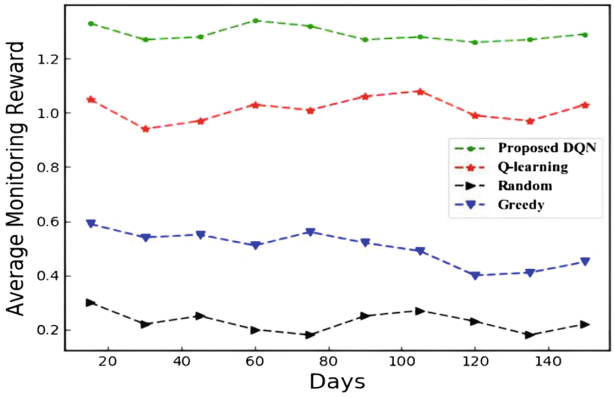
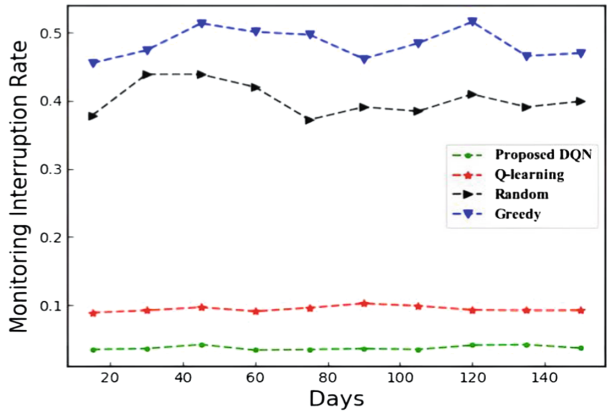


Fig.3. Node state-action diagram (rainy days account for 70%)

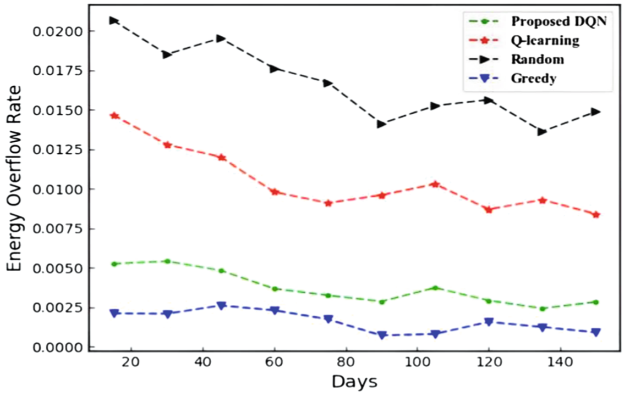
As shown in Fig. 3, we can see that the high probability of rainy days means the harvested energy is relatively lower in the whole 30 days. Compared with scenarios of rainy weather accounted for 50% in the previous subsection, the proposed DQN-based algorithm can also be adapted to the environmental characteristics and further improve



(a)



(b)



(c)

Fig. 4. Performance comparison of Four Algorithms (rainy days account for 70%)

long-term reward of the node by optimized action scheme. At the same time, the residual energy of node can still be maintained into a safe range. Similarly, for the 14th, 15th and 17th sunny days, the proposed algorithm can also adaptively select high monitoring frequency actions to achieve the greater utility. Moreover, in the night-time environment, although nodes cannot harvest energy, the proposed DQN-based algorithm can still select the corresponding action scheme to maintain node monitoring, such that the number of monitoring interruptions is also sharply reduced. It shows that the proposed algorithm can not only improve long-term survival of network nodes, but also improve long-term reward of the whole network. Furthermore, the comparison results of more relevant average monitoring reward, monitoring interruption rate and energy overflow rate are shown in Fig. 4.

From Fig. 4, we can see that the average monitoring reward of proposed DQN-based algorithm is still the greatest among the compared algorithms. Compared with Fig. 4(a), since the harvested energy reduces, the average monitoring reward of proposed algorithm has also been decreased. As shown in Fig. 4(b), although the increase of rainy days makes the available energy of nodes more scarce, the monitoring interruption rate of the proposed algorithm remains below 0.1 and is still the best among the four algorithms. In terms of overflow rate, the proposed algorithm can similarly avoid long-term energy overflow rate by optimized action scheme.

6 Conclusion

In this paper, an adaptive monitoring optimization algorithm based on DQN for EH-WSNs is proposed. We first present the energy harvesting model, energy consumption model and the network model with the single-hop cluster structure. Then, these models are combined with DQN algorithm to improve the monitored performance of the algorithm itself through the mechanism of replay memory and target network. The simulation analysis combines the characteristics of different proportions of rainy weather with day-night alternation to verify the feasibility and effectiveness of the proposed DQN-based algorithm. In addition, we also compare the performance with the other three algorithms under the same simulation environment. The results demonstrate that the proposed DQN-based optimization algorithm can obtain the great performance in terms of three metrics including average monitoring reward, monitoring interruption rate and energy overflow rate. It also indicates that the proposed DQN-based optimization algorithm can not only effectively adapt to the relative complex and changeable weather environment, but also further improve the monitoring utility of network nodes and solve the problem with high monitoring interruption rate of long-term monitoring for EH-WSN in rainy environment.

Acknowledgements. This research was supported by the National Natural Science Foundation of China (Grant No. 61961026, 61962036), Natural Science Foundation of Jiangxi Province, China (Grant No. 20202BABL202003), China Postdoctoral Science Foundation (Grant No. 2020M671556), Major science and technology projects in Jiangxi province (20213AAG01012).

References

1. Lombardo, L., Corbellini, S., Parvis, M., Elsayed, A., Angelini, E., Grassini, S.: Wireless sensor network for distributed environmental monitoring. IEEE Trans. Instrum. Meas. **67**(5), 1214–1222 (2017)
2. Muduli, L., Mishra, D.P., Jana, P.K.: Application of wireless sensor network for environmental monitoring in underground coal mines: a systematic review. J. Netw. Comput. Appl. **106**, 48–67 (2018)
3. Cao, Y., Ji, R., Ji, L., Lei, G., Wang, H., Shao, X.: l^2-MPTCP: a learning-driven latency-aware multipath transport scheme for industrial internet applications. IEEE Transactions on Industrial Informatics (2022)
4. Cao, Y., Ji, R., Huang, X., Lei, G., Shao, X., You, I.: Empirical Mode Decomposition-empowered Network Traffic Anomaly Detection for Secure Multipath TCP Communications, Mobile Networks and Applications (2022)
5. Antony, S.M., Indu, S., Pandey, R.: An efficient solar energy harvesting system for wireless sensor network nodes. J. Inf. Optim. Sci. **41**(1), 39–50 (2020)
6. Sun, W., Ding, Z., Qin, Z., Chu, F., Han, Q.: Wind energy harvesting based on fluttering double-flag type triboelectric nanogenerators. Nano Energy **70**, 104526 (2020)
7. Sharma, H., Haque, A., Jaffery, Z.A.: Modeling and optimisation of a solar energy harvesting system for wireless sensor network nodes. J. Sens. Actuator Netw. **7**(3), 40 (2018)
8. Sharma, H., Haque, A., Jaffery, Z.A.: Maximization of wireless sensor network lifetime using solar energy harvesting for smart agriculture monitoring. Ad Hoc Netw. **94**, 101966 (2019)
9. Sarang, S., Drieberg, M., Awang, A., Ahmad, R.: A QoS MAC protocol for prioritized data in energy harvesting wireless sensor networks. Comput. Netw. **144**, 141–153 (2018)
10. Lakshmi, P.S., Jibukumar, M.G., Neenu, V.S.: Network lifetime enhancement of multi-hop wireless sensor network by RF energy harvesting. In: Proceedings of the 2018 International Conference on Information Networking, pp. 738–743 (2018)
11. Nguyen, H.S., Ly, T.T.H., Nguyen, T.S., Huynh, V.V., Nguyen, T.L., Voznak, M.: Outage performance analysis and SWIPT optimization in energy-harvesting wireless sensor network deploying NOMA. Sensors **19**(3), 613 (2019)
12. Ren, Q., Yao, G.: An energy-efficient cluster head selection scheme for energy-harvesting wireless sensor networks. Sensors **20**(1), 187 (2020)
13. Xiong, Y., Chen, G., Lu, M., Wan, X., Wu, M., She, J.: A two-phase lifetime-enhancing method for hybrid energy-harvesting wireless sensor network. IEEE Sens. J. **20**(4), 1934–1946 (2019)
14. Bengheni, A., Didi, F., Bambrik, I.: EEM-EHWSN: enhanced energy management scheme in energy harvesting wireless sensor networks. Wireless Netw. **25**(6), 3029–3046 (2019)
15. Qiu, C., Hu, Y., Chen, Y., Zeng, B.: Lyapunov optimization for energy harvesting wireless sensor communications. IEEE Internet Things J. **5**(3), 1947–1956 (2018)
16. Lee, P., Eu, Z.A., Han, M., Tan, H.: Empirical modeling of a solar-powered energy harvesting wireless sensor node for time-slotted operation. In: Proceedings of the 2011 IEEE Wireless Communications and Networking Conference, pp. 179–184 (2011)
17. Fraternali, F., Balaji, B., Agarwal, Y., Gupta, R.K.: Aces: automatic configuration of energy harvesting sensors with reinforcement learning. ACM Trans. Sens. Netw. **16**(4), 1–31 (2020)
18. Tekin, N., Gungor, V.C.: The impact of error control schemes on lifetime of energy harvesting wireless sensor networks in industrial environments. Comput. Stand. Interfaces **70**, 103417 (2020)
19. Han, C., Zhang, S., Zhang, B., Zhou, J., Sun, L.: A distributed image compression scheme for energy harvesting wireless multimedia sensor networks. Sensors **20**(3), 667 (2020)
20. Raja, J., Mookhambika, N.: A novel energy harvesting with middle-order weighted probability (EHMoWP) for performance improvement in wireless sensor network (WSN). J. Ambient Intell. Humaniz. Comput. 1–12 (2021)

21. Zairi, S., Zouari, B., Niel, E., Dumitrescu, E.: Nodes self-scheduling approach for maximising wireless sensor network lifetime based on remaining energy. IET Wirel. Sens. Syst. **2**(1), 52–62 (2012)
22. Sahoo, J., Sahoo, B.: Solving target coverage problem in wireless sensor networks using greedy approach. In: Proceedings of the 2020 International Conference on Computer Science, Engineering and Applications, pp. 1–4 (2020)