International Journal of Communication Systems • 2023

Document type
Article

Source type
Journal

ISSN
10745351

DOI
10.1002/dac.5589

Energy-efficient resource allocation over wireless communication systems through deep reinforcement learning

Shukla, Kirti[a]; Kollu, Archana[b]; Panwar, Poonam[c]; Soni, Mukesh[d]; Jindal, Latika[e]; Patel, Hemlata[f]; Keshta, Ismail[g]; Maaliw, Renato R.[h]

Save all to author list

[a] School of Computing Science and Engineering, Galgotias University, Greater Noida, India
[b] Department of Computer Engineering, Pimpri Chinchwad College of Engineering and Research, Pune, India
[c] Faculty of Agriculture, Maharishi Markandeshwar (Deemed to be University), Mullana-Ambala, India
[d] Department of CSE, University Centre for Research & Development, Chandigarh University, Mohali, India

View additional affiliations

Abstract

As the popularity of the Internet of Things (IoT) increases, so do the energy requirements of IoT terminal equipment. To address the energy shortage problem of equipment and ensure continuous and stable operation in light of renewable energy and an uncertain environment, a rational and efficient energy allocation strategy is required. This paper proposes a deep reinforcement learning energy allocation strategy that uses the DQN algorithm to directly interact with the unknown environment. The best energy allocation method is independent of environmental knowledge, and a pretraining algorithm is proposed to maximise the initialization state of the strategy. Experiments of comparison and simulation are conducted under various channel data circumstances. Results indicate that the proposed energy allocation strategy outperforms the current strategy in multiple channel conditions and has a high capacity for adaptation to changing conditions.

Author keywords

deep reinforcement learning; energy harvesting; internet of things; renewable energy; wireless communication system

Sustainable Development Goals 2023
SciVal Topics
Metrics

Indexed keywords

Sustainable Development Goals 2023 New

SciVal Topics

Metrics

Received: 27 February 2023 | Revised: 29 June 2023 | Accepted: 24 July 2023

DOI: 10.1002/dac.5589

RESEARCH ARTICLE

WILEY

Energy-efficient resource allocation over wireless communication systems through deep reinforcement learning

Kirti Shukla[1] | **Archana Kollu**[2] | **Poonam Panwar**[3] | **Mukesh Soni**[4] | **Latika Jindal**[5] | **Hemlata Patel**[6] | **Ismail Keshta**[7] | **Renato R. Maaliw III**[8]

[1]School of Computing Science and Engineering, Galgotias University, Greater Noida, India

[2]Department of Computer Engineering, Pimpri Chinchwad College of Engineering and Research, Pune, India

[3]Faculty of Agriculture, Maharishi Markandeshwar (Deemed to be University), Mullana-Ambala, India

[4]Department of CSE, University Centre for Research & Development, Chandigarh University, Mohali, India

[5]Department of computer science and Engineering, Medicaps University, Madhya Pradesh, India

[6]Computer Science & Engineering, Medicaps University, Madhya Pradesh, India

[7]Computer Science and Information Systems Department, College of Applied Sciences, AlMaarefa University, Riyadh, Saudi Arabia

[8]College of Engineering, Southern Luzon State University, Lucban, Philippines

Correspondence
Mukesh Soni, Department of CSE, University Centre for Research & Development, Chandigarh University, Mohali, Punjab-140413, India.
Email: soni.mukesh15@gmail.com

Summary

As the popularity of the Internet of Things (IoT) increases, so do the energy requirements of IoT terminal equipment. To address the energy shortage problem of equipment and ensure continuous and stable operation in light of renewable energy and an uncertain environment, a rational and efficient energy allocation strategy is required. This paper proposes a deep reinforcement learning energy allocation strategy that uses the DQN algorithm to directly interact with the unknown environment. The best energy allocation method is independent of environmental knowledge, and a pretraining algorithm is proposed to maximise the initialization state of the strategy. Experiments of comparison and simulation are conducted under various channel data circumstances. Results indicate that the proposed energy allocation strategy outperforms the current strategy in multiple channel conditions and has a high capacity for adaptation to changing conditions.

KEYWORDS

deep reinforcement learning, energy harvesting, internet of things, renewable energy, wireless communication system

1 | INTRODUCTION

The term "Internet of Things" (IoT) has been used frequently in recent years, and its application scope has become more and more extensive. However, because the IoT terminal equipment can only carry limited batteries, its energy shortage has always limited the further development of the IoT.[1] Therefore, energy harvesting technology (energy harvesting [EH] is considered to be a promising solution.[2] EH technology is defined as a technology that can collect environmental energy, including, but not limited to, solar, wind, and heat, and transform it into electronic energy.

Therefore, the system with an EH module has some unique advantages: The EH system can continuously collect energy in the case of nonhardware damage, which significantly prolongs the service life of the equipment; it can be deployed in some everyday Hard-to-reach places.[3]

Although the system with the EH module has the above-mentioned remarkable advantages and has been widely used,[4] there is still no suitable dynamic energy allocation strategy to meet the requirements of the EH wireless communication system when the law of change or instability is unknown. Furthermore, under harvesting energy and channel gain[5–8] (the channel gain describes the transmission capability characteristics of the channel itself), the requirements for autonomous work still limit its further large-scale application. To solve the above problems, research currently discusses how to design the optimal access control strategy and energy allocation algorithm for EH-based wireless communication systems.[7–11] Specifically, one idea[7–11] is to design the most efficient means of distributing energy between EH system users and the optimal unloading strategy for MEC systems with EH modules. Another idea is to use the Markov decision process-based dynamic programming algorithm,[12–15] that is, by using the system's power distribution issue is modelled using a Markov decision procedure and then using the dynamic programming method to obtain the optimal power allocation scheme. The above techniques all rely directly or indirectly on the collection of priority and the system knowledge of energy distribution and channel gain distribution. This prior knowledge is complicated to obtain in actual use. Even if the random distribution corresponding to the environmental change in a period is obtained through sampling, the random distribution is in the subsequent time. There is also the possibility of continuous changes, making the final model unable to adapt to the unknown environment.

Therefore, given the difficulty of solving the problem of prior knowledge, people turn to some model-free learning-based methods to reduce or even eliminate the constraints of previous knowledge. Among them, reinforcement learning is a method that allows agents to learn autonomously in an unknown environment. Considering that the characteristics of reinforcement learning are very suitable for the parts of EH systems, this paper gives some examples of the application of reinforcement learning in EH systems. In literature,[16] reinforcement learning the Q-learning method in the article is applied to a two-hop EH relay system to maximise the system throughput. In Shukla et al.,[17] the Q-learning system is used for an EH sensor node so that the node can ensure energy storage while maximising the sampling rate. In the literature,[18] the EH wireless sensor network (WSN) uses the Q-learning system to achieve the ideal data packet delivery rate under limited prior knowledge. The literature[18,19] adopts the deep reinforcement of actor-critic. The transmission scheme of the wireless communication system based on EH is designed using the learning method. The literature[19] considers the multi-access system in which the access control access point (AP) based on the EH module transmits data to multiple users and uses the deep reinforcement learning method to design the energy distribution and multiple-access control policies. These findings demonstrate that reinforcement learning can potentially address policy issues in EH systems in unknown environments.

The objective of this paper is to address the energy shortage problem in IoT terminal equipment by proposing a DQN-based reinforcement learning energy allocation strategy for EH multiple-access wireless communication systems in an unknown environment. The proposed method aims to achieve online rational planning and control of APs to select multiple users to access various channels, with the goal of joint collaborative optimization to maximise the system's long-term throughput and working time. The paper also introduces an improved algorithm for prelearning based on reinforcement learning and the time-invariant structure characteristics of the system to improve the system's variable scene learning ability and learning performance in the early stage. The experiments demonstrate that the proposed method yields superior outcomes than the traditional strategy in simulation experiments, and the pretraining algorithm optimises the system's initial state with apparent effect, providing better learning ability in changing scenarios.

Despite the widespread use of EH modules in IoT systems, there is still a lack of suitable dynamic energy allocation strategies to address the requirements of EH wireless communication systems in unknown or changing environments. Existing research has focused on designing optimal access control strategies and energy allocation algorithms for EH-based wireless communication systems. However, these approaches often rely on prior knowledge of energy distribution and channel gain, which is challenging to obtain and may not be applicable in real-world scenarios. The need for efficient energy allocation strategies that can adapt to unknown environments and eliminate prior knowledge constraints remains unmet.

Additionally, while reinforcement learning has shown promise in addressing policy issues in EH systems, there is a need for further exploration of its application in unknown environments. Previous studies have applied reinforcement learning, such as Q-learning and deep reinforcement learning, in EH systems with promising results. However, more research is required to investigate the potential of reinforcement learning methods, particularly in the context of EH multiple-access wireless communication systems. Specifically, there is a research gap in developing reinforcement

learning-based energy allocation strategies that can achieve online rational planning and control of APs, maximise long-term throughput, and improve the system's learning ability in changing scenarios.

Contribution of the work is as follows:

This paper addresses the energy shortage problem in IoT terminal equipment by proposing a novel DQN-based reinforcement learning energy allocation strategy for EH multiple-access wireless communication systems in unknown environments.

1. The proposed method focuses on achieving online rational planning and control of APs, enabling the selection of multiple users to access different channels. The objective is to conduct joint collaborative optimization, maximising the system's long-term throughput and working time.
2. Furthermore, the paper introduces an improved algorithm for prelearning based on reinforcement learning and the time-invariant structure characteristics of the system. This algorithm enhances the system's learning performance and variable scene learning ability, particularly in the early stages when facing changing scenarios.
3. The proposed method demonstrates superior outcomes through simulation experiments compared to traditional strategies. The pretraining algorithm optimises the system's initial state effectively, enhancing learning ability in dynamic scenarios.
4. The contributions of this work include the development of a DQN-based reinforcement learning energy allocation strategy for EH multiple-access wireless communication systems without relying on prior system knowledge.
5. Additionally, introducing the improved prelearning algorithm enhances the system's learning performance and adaptability to variable scenes, ultimately improving the overall efficiency of EH systems in unknown environments.

2 | RELATED WORK

In this overview of the literature, we examine a number of studies on resource allocation methods based on deep reinforcement learning in diverse fields. The proper distribution of resources is essential for maximising the performance of a variety of systems, from wireless networks to industrial settings. Deep reinforcement learning techniques have recently come to light as viable solutions to the problems associated with effective resource allocation. This review of the literature provides an overview of recent studies that use deep reinforcement learning to allocate resources and highlights their contributions to the area.

Deep reinforcement learning was suggested as a better method of communication resource allocation for wireless networks by Ting Xu et al.[20] They used the capability of deep reinforcement learning techniques to address the problem of resource allocation optimisation. The authors showed that their method improved the overall performance of wireless networks.

Ying Wang et al.[21] introduced a dual-attention assisted deep reinforcement learning system for energy-efficient resource allocation in the Industrial Internet of Things (IIoT) scenario. Their strategy improved resource allocation in IIoT systems by using dual-attention methods. By taking into account the particular needs and difficulties of resource allocation in IIoT contexts, this work made a contribution to the field.

A multigranularity fusion resource allocation approach for heterogeneous IIoT was put forth by Ying Wang et al.[22] and is based on dual-attention deep reinforcement learning and lifelong learning architecture. The complexity of resource allocation in various IIoT contexts was solved by their algorithm by merging dual-attention deep reinforcement learning and lifelong learning. The authors emphasised the significance of taking into account various granularity levels when deciding how to allocate resources.

A cooperative deep reinforcement learning approach was used by Ying Wang et al.[23] to develop a routing algorithm for software-defined WSNs that is both energy-efficient and delay-guaranteed. Their work concentrated on routing choice optimisation while taking energy usage and delay requirements into account. The authors showed how their strategy improved the performance of software-defined WSNs by utilising deep reinforcement learning techniques.

A deep reinforcement learning-based resource optimisation technique for UAV-assisted mobile edge computing (MEC) systems was put forth by Fan Yu et al.[24] The difficulties of effective resource allocation in the context of UAVs and edge computing were solved by their method. By employing deep reinforcement learning, the authors attempted to maximise resource utilisation in UAV-assisted MEC systems.

A deep transfer reinforcement learning method for resource allocation in hybrid, multiple-access systems was introduced by Xiaoming Wang et al.[25] Their research centred on how to allocate resources more effectively in hybrid

multiple-access systems. The authors sought to increase the effectiveness of resource allocation in such situations by utilising deep transfer reinforcement learning.

For heterogeneous Vehicle-to-Everything (V2X) networks, Junhui Zhao et al.[26] suggested a multi-agent deep reinforcement learning-based resource management solution. Their research addressed problems with resource allocation in V2X networks and showed how multi-agent deep reinforcement learning can optimise resource allocation choices.

In a 6G-enabled edge environment powered by Cybertwin, Vibha Jain et al.[27] presented a resource allocation strategy employing deep reinforcement learning. To optimise resource distribution in 6G-enabled edge contexts, they used deep reinforcement learning algorithms. The advantages of deploying a Cybertwin for effective resource management were emphasised by the writers.

For distributed AI execution tasks in IoT edge computing environments, Zahra Aghapour et al.[28] suggested a task offloading and resource allocation technique based on deep reinforcement learning. They optimised judgements about task offloading and resource allocation in IoT edge computing systems. The benefits of deep reinforcement learning for effective resource utilisation were underlined by the authors.

A reinforcement learning-based strategy for flow and energy management in resource-constrained wireless networks was presented by Hrishikesh Dutta et al.[29] The difficulties with managing energy and flow in wireless networks with limited resources were the main subject of their study. The authors' method for optimising resource use was successfully proven through the use of reinforcement learning techniques.

In nonorthogonal multiple access (NOMA) systems, Supraja G. et al.[30] suggested maximising throughput and providing dependable wireless communication. Their research focused on resource allocation techniques in NOMA systems for optimum dependability and throughput. The benefits of using a linked fog structure and a weighted energy efficiency power allocation approach were emphasised by the authors.

Taking into account the effects of energy storage losses, Abdul Basit et al.[31] presented an optimal power allocation approach for energy-harvesting wireless communication systems. With consideration for energy storage losses, their strategy intended to optimise power allocation choices in energy-harvesting wireless communication systems. The authors showed that their approach improved the system's overall performance.

For 6G wireless systems with network slicing, Jie Huang et al.[32] presented an opportunistic capacity-based resource allocation technique. In their work, which addressed resource management issues in 6G wireless systems, they emphasised the advantages of an opportunistic capacity-based resource allocation technique for maximising resource utilisation.

A joint task processing/offloading mode selection and resource allocation method for backscatter-aided and wireless-powered MEC systems was presented by KunLi Shi et al.[33] Their research focused on optimising resource allocation and task processing/offloading choices in MEC systems powered by wireless and backscatter. The benefits of the authors' cooperative optimisation strategy were emphasised.

An enhanced deep reinforcement learning-based communication resource allocation technique for wireless networks was put forth by Ting Xu et al.[20] Their study addressed the issues with wireless network resource allocation and showed how deep reinforcement learning techniques may be used to optimise communication resource allocation decisions.

A joint trajectory-resource optimisation method for UAV-enabled uplink communication networks with wireless backhaul was introduced by Bo Hu et al.[34] In UAV-enabled uplink communication networks, their work concentrated on optimising trajectory and resource allocation selections. The authors demonstrated how their strategy helped them achieve effective resource utilisation and enhance network performance.

A comparison of recent publications and review papers in the area of deep reinforcement learning (RL) used in wireless communication systems is shown in the table below. These studies investigate how deep RL approaches can be used to manage and allocate resources more efficiently in diverse wireless network environments. The goal of article summary Table 1 is to provide a brief summary of each work's main characteristics, contributions, and limitations as well as an overview of the relevant fields of research.

Although the presented studies have made significant contributions to the resource allocation field using deep reinforcement learning, research gaps still warrant further investigation. Specifically, there is a need for more studies that explore the application of deep reinforcement learning in resource allocation for specific domains such as industrial environments, IoT edge computing, and heterogeneous networks. Additionally, developing more efficient and scalable algorithms for resource allocation is an area that requires further attention. By addressing these research gaps, researchers can continue to advance the resource allocation field and pave the way for more efficient and effective resource management in diverse systems.

TABLE 1 Article summary.

Paper reference	Key features and contributions	Focus/coverage	Limitations
Ting Xu et al[20]	Improved resource allocation strategy using deep RL for wireless networks.	Wireless networks	No focus on EH multiple-access systems.
Ying Wang et al.[21]	Dual-attention deep RL algorithm for resource allocation in IIoT.	IIoT resource allocation	Limited to IIoT context, no focus on EH multiple-access systems.
Ying Wang et al.[22]	Multigranularity fusion resource allocation algorithm based on dual-attention deep RL in heterogeneous IIoT.	Heterogeneous IIoT resource allocation	Limited to IIoT context, no focus on EH multiple-access systems.
Ying Wang et al.[23]	Energy-efficient and delay-guaranteed routing algorithm using cooperative deep RL for software-defined WSNs.	Software-defined WSN routing optimization	Limited to routing optimization, no focus on EH multiple-access systems.
Fan Yu et al.[24]	Resource optimization algorithm for UAV-assisted mobile edge computing systems based on deep RL.	UAV-assisted mobile edge computing resource allocation	No focus on EH multiple-access systems, limited to UAV and edge computing context.
Xiaoming Wang et al.[25]	Deep transfer RL approach for resource allocation in hybrid multiple-access systems.	Hybrid multiple-access resource allocation	No focus on EH systems, limited to hybrid multiple-access scenarios.
Junhui Zhao et al.[26]	Multi-agent deep RL-based resource management approach for heterogeneous V2X networks.	Heterogeneous V2X resource allocation	No focus on EH multiple-access systems, limited to V2X networks.
Vibha Jain et al.[27]	Resource allocation method using deep RL in 6G-enabled edge environments driven by Cybertwin.	6G-enabled edge resource allocation	No focus on EH multiple-access systems, limited to 6G-enabled edge environments.
Zahra Aghapour et al.[28]	Task offloading and resource allocation algorithm based on deep RL for distributed AI execution tasks in IoT edge computing.	IoT edge computing resource allocation and task offloading	No focus on EH multiple-access systems, limited to IoT edge computing.
Hrishikesh Dutta et al.[29]	RL-based approach for flow and energy management in resource-constrained wireless networks.	Resource-constrained wireless network management	No focus on EH multiple-access systems, limited to resource-constrained wireless networks.
Supraja G. et al.[30]	Resource allocation strategies in NOMA systems for maximising throughput and reliability.	NOMA resource allocation	No focus on EH multiple-access systems, limited to NOMA systems.
Abdul Basit et al.[31]	Optimum power allocation method for EH wireless communication systems considering energy storage losses.	EH power allocation	Focus on power allocation, limited to EH wireless communication systems.
Jie Huang et al.[32]	Opportunistic capacity-based resource allocation approach for 6G wireless systems with network slicing.	6G wireless resource allocation	No focus on EH multiple-access systems, limited to 6G wireless systems.
KunLi Shi et al.[33]	Joint task processing/offloading mode selection and resource allocation in backscatter-aided and wireless-powered MEC systems.	Backscatter-aided and wireless-powered MEC resource allocation	No focus on EH multiple-access systems, limited to MEC systems.
Ting Xu et al.[20]	Improved communication resource allocation strategy for wireless networks based on deep RL.	Wireless network resource allocation	No specific focus on EH multiple-access systems.
Bo Hu et al.[34]	Joint trajectory-resource optimization approach for UAV-enabled uplink communication networks with wireless backhaul.	UAV-enabled uplink network optimization	No focus on EH multiple-access systems, limited to UAV-enabled uplink networks.

Abbreviations: EH, energy harvesting; IIoT, Industrial Internet of Things; MEC, mobile edge computing; RL, reinforcement learning; WSN, wireless sensor network.

3 | SYSTEM MODEL

In a multichannel wireless communication model with EH, devices collect energy from the environment and use it to power their communication transmissions. The devices are equipped with multiple channels to transmit information, and each channel has different characteristics, such as noise level, interference, and channel gain. The EH process can be intermittent and unpredictable, adding complexity to the system's operation. To optimise the use of energy resources, efficient resource allocation strategies must be designed to allocate energy and select the best channel for transmission based on the current environment's conditions. This is typically done using machine learning algorithms, such as deep reinforcement learning, to learn from past experiences and improve future decision-making. By developing efficient energy allocation strategies, devices can prolong their battery life and operate in energy-constrained environments for longer periods (Figure 1).

3.1 | Channel model

The system shown in Figure 2 is a typical multichannel wireless communication model with EH, which consists of an EH module, an access control AP, a set of u user points and a set of k OFDM (orthogonal frequency division multiplexing) channel composition.

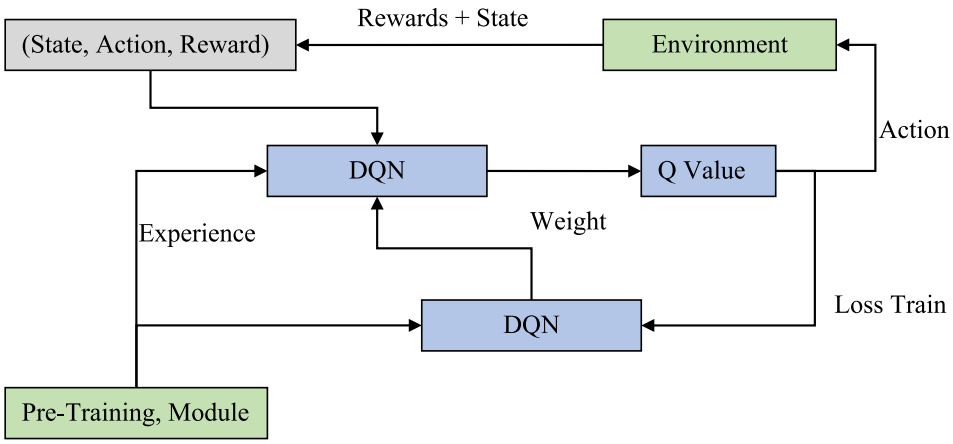

FIGURE 1 Graphical abstract for the proposed methodology.

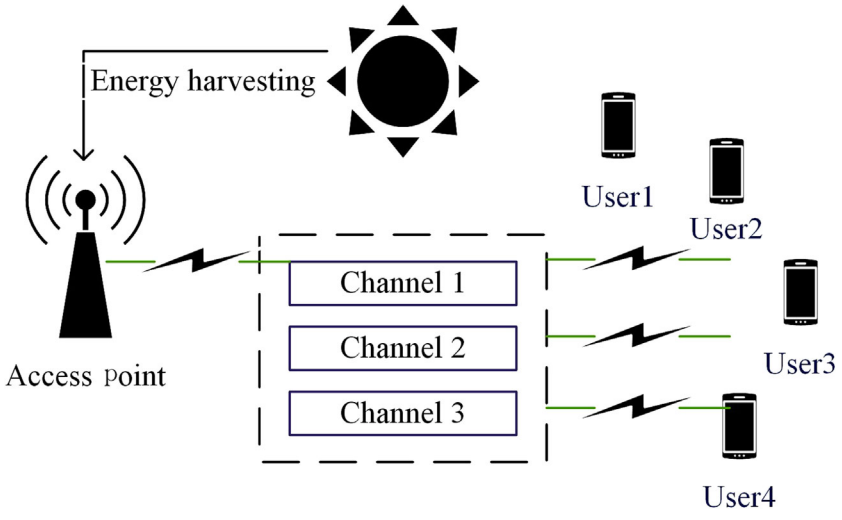

FIGURE 2 Multichannel wireless communication model with energy harvesting.

Only a few discrete communication models, matching to various channel coding rates, are currently enabled by the standard wireless communication system and modulation sequences. Different transmission data rates have additional requirements for received signal power. Usually, more A high transmission data rate corresponds to an ample received signal power. In this paper, R = $\{r_1, r_2, \cdots, r_M\}$ represents the transmission rate of M files, and T_M describes the smallest received signal power corresponding to the transmission rate r_M.

The wireless system works in the form of time slots, so it is natural to record the channel gain of the i-th channel corresponding to the kth user in the t-th time slot as $C_k^i[t]$ (the channel gain here refers to the transmission energy consumption corresponding attenuation factor), and C[t] is a matrix with a dimension of n × k, which represents the situation that all user points correspond to all channel numbers. Considering that the size of the time slot is usually small, we assume that in a single time slot, the channel gain in the slot is constant; that is, the corresponding channel gain in each time slot is an independent and identically distributed random variable, which obeys the highly correlated random distribution of the environmental model. The channel gain in the entire time slot sequence is time-varying. At the beginning of each time slot, there are many ways to obtain the channel gain in the current time slot. The more typical method is to use a pulse signal; that is, the user sends a short pulse signal to the AP with fixed power, and the incoming pulse the signal strength estimates the current instantaneous channel gain. This paper assumes that the pulse signal uses a static transmission power P to transmit the signal, so to provision a transmission rate operation, the channel gain needs to satisfy $P\boldsymbol{C}_k^i[t] > T_M and P\boldsymbol{C}_k^i[t] \leqslant T_{M+1}$; thus,

$$T_M/P < \boldsymbol{C}_k^i[t] < T_{M+1}/P \tag{1}$$

From Formula (1), it can be seen that the channel quality $\boldsymbol{C}_k^i[t]$ can be restrained entirely by the smallest received signal power TM. Then the channel gain state can be measured by the data transmission rate. However, it should be noted here the worst data transmission. Therefore, the rate cannot be provided, and this paper chooses to record the data transmission rate that cannot be supplied as $r_0 = 0$. Therefore, the data transmission rate is directly used to describe the channel gain in the following. Further, the variance corresponds to the distribution that the data transmission rate obeys and the mean can be used to measure the quality of the overall channel state.

3.2 | EH module model and battery model

This article assumes that the EH module is always in working condition but also considers the working situation without the EH module under exceptional circumstances. Each time slot and the interval between time slots in the system work are the exact sizes. For the sake of simplicity, it is considered that each time slot is the interval between the time slot and the previous time slot of this time slot; that is, the energy collected during this period is the energy contained in this time slot and is recorded as E[t]. The battery has an upper capacity limit of B_{max}, and each time slot starts, the energy at a time is recorded as B[t]. Specifically, E[t] and B[t] are discretized in this paper, and unit 1 is the energy consumption of transmitting data packets. This is because the scale of the system is generally the same as matched; ideally, the energy collected in each time slot can support the data transmission action of the current round.

Much like other studies,[16] we view the entire cycle of acquiring and using energy as a Markov decision process. In each time interval, we keep track of the power consumption (P[t]) necessary to complete the task. Calculate the working energy P [t] at the conclusion of each time window and the collected energy E [t] at the commencement of each time interval. The Markov decision procedure can be used to predict the value of the energy E[t] in the following time interval [t + 1].

$$B[t+1] = min\{B_{max}, E[t] + B[t] - P[t]\} \tag{2}$$

3.3 | System workflow

This paper assumes that all signal transmissions use a binary transmission strategy; there are only two possibilities of providing complete service and no service, the same as the system used in the literature.[2,19] In each time slot, a user only one channel can be allocated at most, and each channel can only be used for at most one user to transmit information.

Like the channel state matrix, we write the sending action in the time slot t, the channel allocation, as a matrix A[t] of the same size as C[t]. The element $\boldsymbol{A}_k^i[t]$ in A[t] satisfies $\boldsymbol{A}_k^i[t] \in \{0,1\}$, where $\boldsymbol{A}_k^i[t] = 0$ means that the i-th channel is not assigned to the kth user, $\boldsymbol{A}_k^i[t] = 1$ means allocation. Then the natural total transmission power is as follows:

$$P[t] = \left(\sum_n \sum_k \boldsymbol{A}_k^n[t] \right) P \tag{3}$$

Among them, P is the fixed transmission power.

If there is not enough power in the system to carry out the transmitting action chosen for this round (i.e., P[t] > B [t]) at the start of the time period, the system will skip that action and move on to the next. Since the processing power initiated by the EH circuit is typically miniscule in comparison with the real transmission power, we presume that the power needed to activate the AP's EH circuit is similarly insignificant.

4 | PROBLEM FORMULATION AND DEEP Q-LEARNING FRAMEWORK

By allowing an agent to learn from its own mistakes in a given scenario, reinforcement learning can model the agent's interactions with its surroundings and make the best possible decisions. An agent gets information about the current condition of the world at each time step and chooses an appropriate course of action in reaction. The agent then receives the incentive value and the new condition based on the input from the world in order to continue learning the intended policy. Figure 3 depicts the precise method being used. However, before reinforcement learning can be used, the world must be modelled as a Markov decision process.

4.1 | Form of Markov decision process

Markov decision procedure can usually be represented by a quadruple (s, a, r, T): state space s, action space a, reward function r and state transition function T. However, under the unknown environmental conditions considered, giving a definite state transition function T is impossible, so only the state space s, action space a and reward function r are given below.

1. State space s[t]: It consists of the channel state space (C[t], E[t]) of the current time slot, namely:

$$\mathrm{s}[\mathrm{t}] \in \mathrm{S} = \{\mathrm{C}[\mathrm{t}], \mathrm{B}[\mathrm{t}], \mathrm{E}[\mathrm{t}]\}$$

2. The action space a[t] and C[t] have the same latitude; use 0, 1 to indicate which channel to choose for transmission.
3. Reward function r[t]:

$$r(s[t], a[t]) = \begin{cases} \alpha \sum_n \sum_k (a[t] \circ s[t])_n^k, & P[t] \leqslant B[t] \\ -\beta, & P[t] > B[t] \end{cases}$$

The limitation of the traditional Bellman equation iteration method is that it depends on the determined state transition function T.

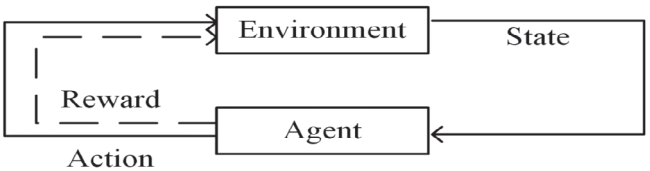

FIGURE 3 Schematic diagram of reinforcement learning process.

4.2 | Deep reinforcement learning

Deep reinforcement learning (DRL) has several advantages for designing energy-efficient resource allocation over wireless communication systems:

- **Flexibility:** DRL is a flexible approach that can adapt to different scenarios, environments and system requirements without the need for a priori knowledge. This flexibility allows DRL to optimise resource allocation in dynamic and uncertain wireless communication systems.
- **Energy efficiency:** DRL can optimise resource allocation to reduce energy consumption in wireless communication systems. By reducing energy consumption, DRL can extend the battery life of IoT devices and reduce the carbon footprint of wireless communication systems.
- **Learning ability:** DRL can learn from experience and improve its performance over time. This learning ability allows DRL to adapt to changing wireless communication environments, user behaviour and system requirements.
- **Scalability:** DRL can scale to large and complex wireless communication systems with many devices, channels and users. This scalability allows DRL to optimise resource allocation in large-scale wireless communication networks, such as smart cities and industrial IoT.
- **Performance:** DRL can achieve better performance than traditional optimization methods in wireless communication systems. By optimising resource allocation based on the current state and future rewards, DRL can achieve higher throughput, lower delay and better Quality of Service (QoS) in wireless communication systems.

4.2.1 | Q-learning

The core concept in the Q-learning algorithm is to evaluate the quality of acting on the state S through the action value function Q(s, a). To maximise the long-term average reward, that is, to obtain some of the best actions, the Q-learning algorithm uses experience (s[t], a[t], r[t], s[t + 1]) to learn the action value function Q(s, a). Specifically, use the formula in the time difference method:

$$Q(s,a) \leftarrow Q(s,a) + \alpha\left[r + \gamma \underset{a'}{max} Q(s',a') - Q(s,a)\right] \tag{4}$$

Execute the above formula recursively on all experience sets until convergence, where $\alpha \in (0,1)$ is the learning rate. After all Q(s, a) are generated, the optimal strategy—the set of optimal actions—is a crucial step of Q-learning to generate experience (s[t], a[t], r[t], s[t + 1]), that is, learning samples. The usual method of generation is the classic ε-greedy algorithm:

$$a = \begin{cases} \arg\underset{a}{max} Q(s,a), \text{ with probability } \varepsilon \\ \text{randomly chose another action} \end{cases} \tag{5}$$

The ε-greedy algorithm "uses" the current best action with a certain probability and then "explores" unselected efforts with a certain chance. Outdated Q-learning uses a Q table to store Q(s, a) with (S, A) as the row and column. Still, When the scale of Q(s, a) is large, the sparse Q table not only consumes a lot of storage space but also causes inefficiency in exploration. Therefore, with the progress of deep learning in recent years, reinforcement learning naturally introduces neural networks to overcome the difficulty of storing large Q-tables.

4.2.2 | Deep Q-learning

The idea of deep Q-learning (DeepQ-network, DQN) is to use a set of weights to approximate the representation of Q(s, a), that is, $Q(s, a, \theta) \simeq Q(s, a)$.

The purpose of learning to represent Q(s, a) with Q(s, a,θ) is achieved by minimising the appropriate loss function $L_i(\theta_i)$.

$$L_i(\theta_i) = \sum\nolimits_D \mathrm{E}\left[\left(R + \gamma \max_{a'} Q(s', a'; \theta_{i-1}) - Q(s, a; \theta_i)\right)^2\right] \tag{6}$$

On this basis, add experience pool playback and double neural network (DoubleDQN, DDQN) staggered update technology to DQN to overcome the difficulty of neural network training caused by the high correlation and unbalanced distribution of learning samples. Experience pool playback refers to all pieces being put into a fixed-size experience pool. Each time a new model is generated, the experience pool is updated, and each training takes a fixed number of batches of samples from the experience pool. A double neural network uses two settings that are precisely the same. Each time one of them generates an action and trains to update the other network, and after a sureamount of rounds, the parameters of this neural network are thoroughly copied to the other. Therefore, DQN and DDQN essentially refer to the same, and the following algorithm framework at this time is shown in Figure 4. Even if the above two techniques are used, the training time of the neural network is still relatively long.

4.3 | Pretraining module and overall process

When the system uses changing scenarios, we all hope to speed up the learning process to reach the state of using the optimal strategy in the shortest time. Some characteristics of the considered system are independent of environmental factors, such as channel gain at each time C[t], the total system initial energy E[t], the scale of the system includes the number of orthogonal channels and the extreme number of users.

Specifically, the action selection is based on the existing energy E[t] and the present channel state C[t]. Suppose the existing energy E[t] stored in the battery is a relatively small value (almost empty). In that case, the system tends to adopt a more conservative strategy (that is, select as few or no users as possible for communication). Vice versa, if the energy E[t] stored in the battery is almost complete, the system often adopts a more active strategy. Furthermore, if a user is selected to transmit data, not all channels can be allocated; only the medium with the best channel status may be selected, which is not difficult. It is extended to the combination of multiple channels. Therefore, the action selection of a strategy should be divided into two parts: action selection propensity and specific action selection. These two parts are the parts that the system can learn, and all the parts that the public can know are included in the learning information in the form of hidden knowledge and entered into the Q value.

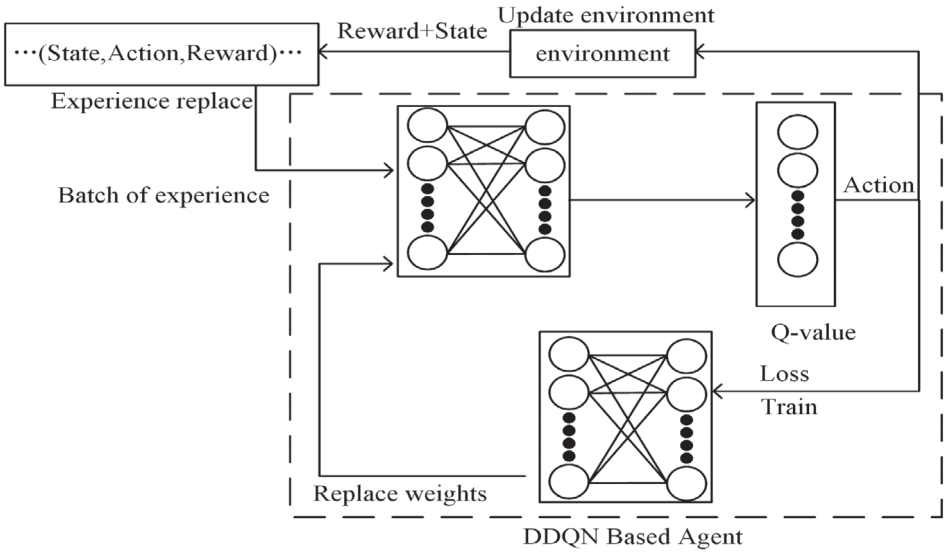

FIGURE 4 Algorithm framework based on DoubleDQN (DDQN).

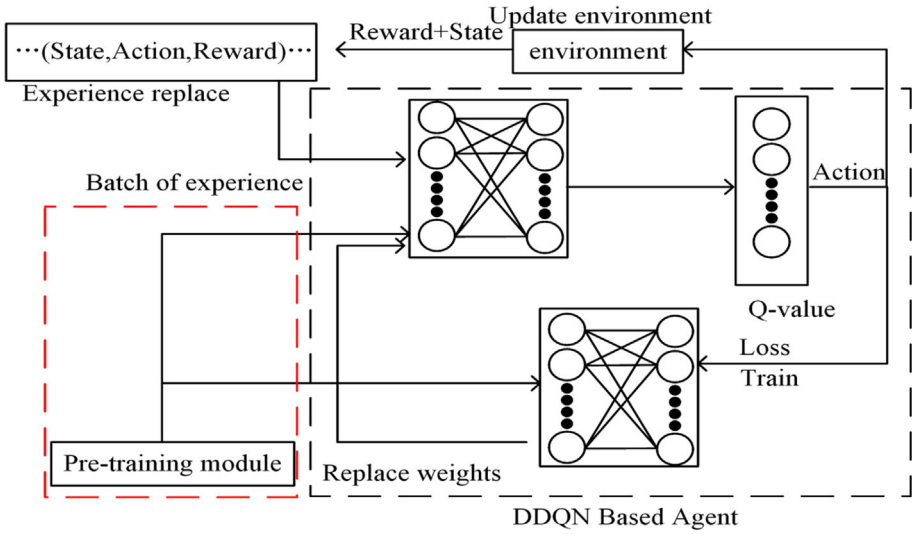

FIGURE 5 Algorithm framework based on DoubleDQN (DDQN) with pretraining modules (electronic version is in colour).

Based on the above understanding, no matter which environment model is used, the system can learn some prior structural features in advance through transfer learning from the learned Q(s, a) part. Therefore, we add a pretraining module and assign a more appropriate preset initial Q value Q(s, a) for DQN. We get a set of neural network weights and build them into the system by letting the method learn more generally. Representative hypothetical environment In, assign an appropriate initial Q value Q(s, a) to our target DQN to reduce the overhead of unknown action exploration so that the system can learn an optimal strategy faster. To sum up, the overall framework of the algorithm in this paper is shown in Figure 5, and the part of the pretraining module is marked with a red box in Figure 5.

5 | SIMULATION EXPERIMENT

The setup for simulation experiments to test the performance of the proposed algorithm and verify the results is described in this section. The main evaluation index used to assess the algorithm's performance is the (average) long-standing throughput of the system, which measures how much data can be transmitted over a period of time. The experimental environment where the simulations are run is specified as Window10, and the programming language used to implement the algorithm is mentioned as Python3.7. The deep learning framework used for the implementation is also noted to be Keras. It is worth noting that Keras is a high-level neural network API written in Python and is capable of running on top of TensorFlow, CNTK or Theano. The process of building deep learning models is simplified by Keras by providing a user-friendly API that allows developers to quickly prototype and experiment with various architectures. By specifying the experimental environment, programming language and deep learning framework used, the replicability of the simulation experiments by other researchers is ensured, which is essential for building upon existing work and advancing the field.

5.1 | Simulation experiment setup

The system parameter settings of the experimental test are listed in Table 2.

Considering that the scale of actions corresponding to the experimental system is relatively massive, $|\boldsymbol{A}| = \sum_{i=0}^{4} \boldsymbol{C}_6^i \times \boldsymbol{A}_6^i = 1045$, so the experiment uses a rather sizeable neural network scale, and the specific parameter settings of the neural network are listed in Table 3. The parameters in the reward function are set to α, β; that is, the reward function considered in this paper is as follows:

$$R(S[t],\boldsymbol{A}[t]) = \begin{cases} \sum_{n}\sum_{k} (\boldsymbol{A}[t] \circ S[t])_n^k, & P[t] \leqslant B[t] \\ -2, & P[t] > B[t] \end{cases}$$

TABLE 2 System parameter settings.

Parameter	Value
Number of users	6
Number of channels	4
Limitation of battery	35\70
Data rate	[0,1,2,3]
EH energy	[1,3,5,7]
The parameters of the reward function	$\alpha = 1.0, \beta = 2.0$

TABLE 3 Neural network parameter settings.

Parameter	Value
Number of hidden layer	3
Number of hidden layer nodes	260

TABLE 4 Test environment settings.

Environment name	Distribution of channels	Channel probability	Distribution of EH	EH probability	Initial energy
EnV1	Gaussian ($\mu = 2, \sigma = 1$)	[0.138,0.356,0.361,0.145]	-	-	75
EnV2	Gaussian ($\mu = 4, \sigma = 1$)	[0.029,0.042,0.282,0.678]	-	-	75
EnV3	Gaussian ($\mu = 2, \sigma = 1$)	[0.139.0.357,0.360,0.144]	Poisson ($a = 0.85$)	[0.455,0.357,0.143,0.047]	35
EnV4	Gaussian ($\mu = 4, \sigma = 1$)	[0.029,0.041.0.280,0.676]	Poisson ($\lambda = 0.85$)	[0.455,0.357,0.143,0.047]	35
EnV5	Gaussian ($\mu = 2, \sigma = 1$)	[0.139,0.357,0.360,0.144]	Poisson ($\lambda = 1.9$)	[0.145,0.291,0.282,0.281]	35
EnV6	Gaussian ($\mu = 4, \sigma = 1$)	[0.029,0.041.0.280.0.676]	Poisson ($\lambda = 1.9$)	[0.145,0.291.0.282,0.281]	35
EnV7	Uniform	[0.250,0.250,0.250,0.250]	Uniform	[0.250,0.250,0.250,0.250]	35

Abbreviation: EH, energy harvesting.

This paper selects the most typical six environmental models and one preset model in the usage scenario, among which the environmental models Env1-Env2 correspond to the unique working conditions without the EH module; Env3-Env6 correspond to the normal working conditions under different ecological conditions and channel conditions; Env7 is the environment model used for pretraining. The corresponding probability value is obtained by sampling the corresponding distribution, as listed in Table 4.

For comparative experiments, some traditional representative strategies are selected:

1. Round-robin strategy (round-robin). The 2G polling strategy adopted in this paper is a simple load-balancing algorithm. It follows a fixed order based on the user sent last time and sequentially sends services to the following i users. The approach does not take into account any overhead or long-term cumulative returns.
2. Random strategy (random). This strategy chooses an arbitrary action to execute, a spontaneous movement in an action space.
3. Optimal strategy (optimal). This is also an instant optimal approach, which always chooses the action that exploits the immediate reward in the recent time slot. In other words, in states, it always determines the following steps:

$$a = \arg \max_{\forall \boldsymbol{A} \in \boldsymbol{n}_{dl_s}} \left\{ \sum_n \sum_k (\boldsymbol{A} \circ S[t])_n^k \right\}$$

To correctly evaluate the pros and cons of the algorithm, this paper sets the following three performance indicators as the criteria for assessing the pros and cons of the algorithm.

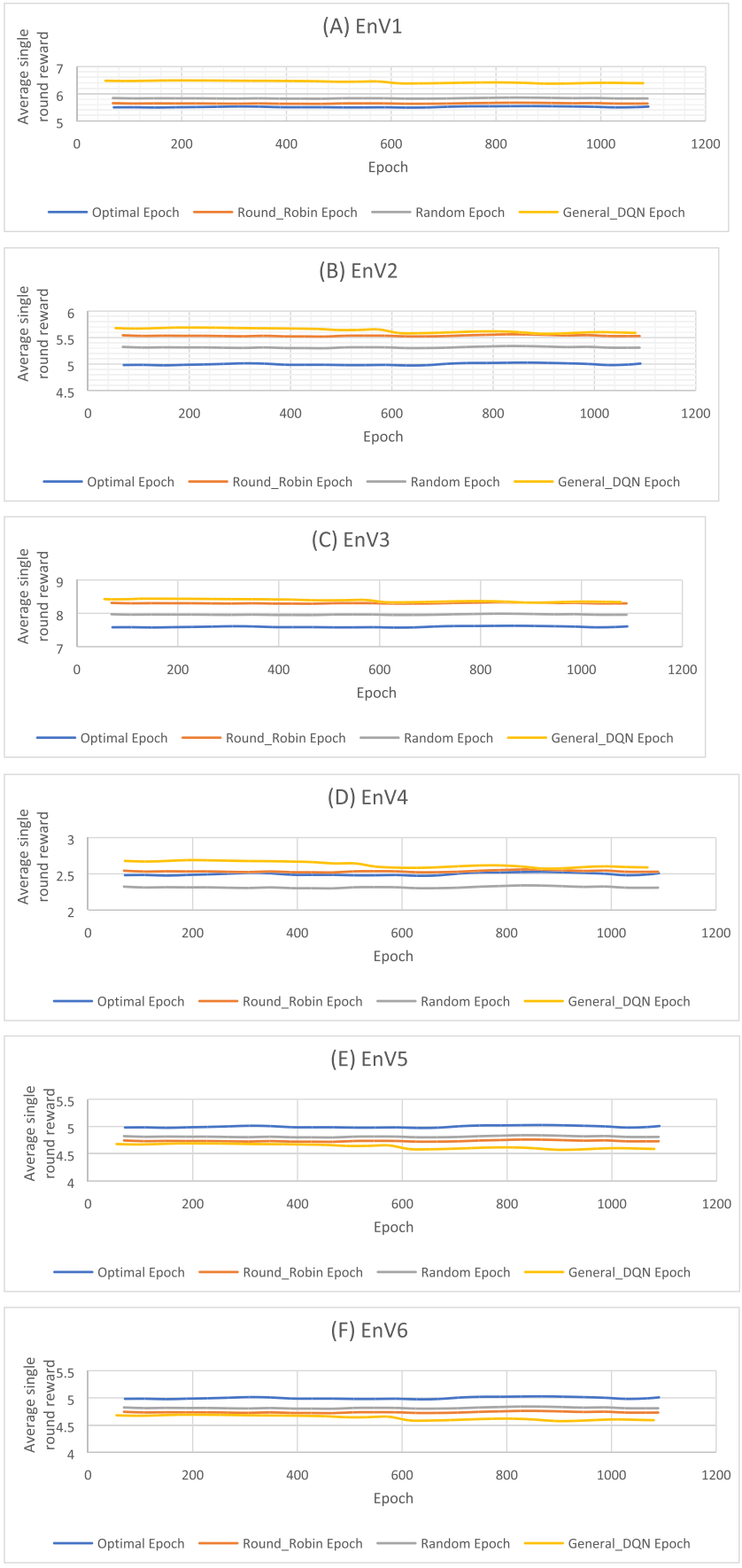

FIGURE 6 Performance comparison in various environments.

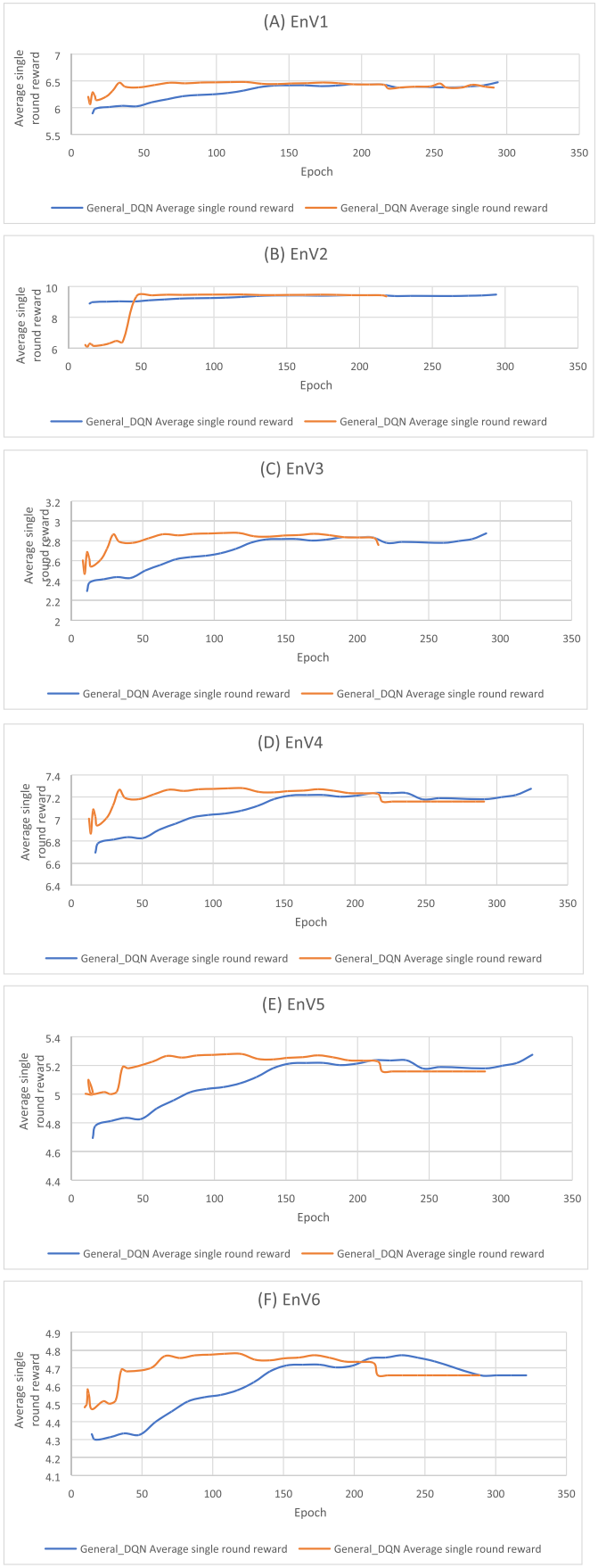

FIGURE 7 Comparison of pretraining acceleration effects in various environments.

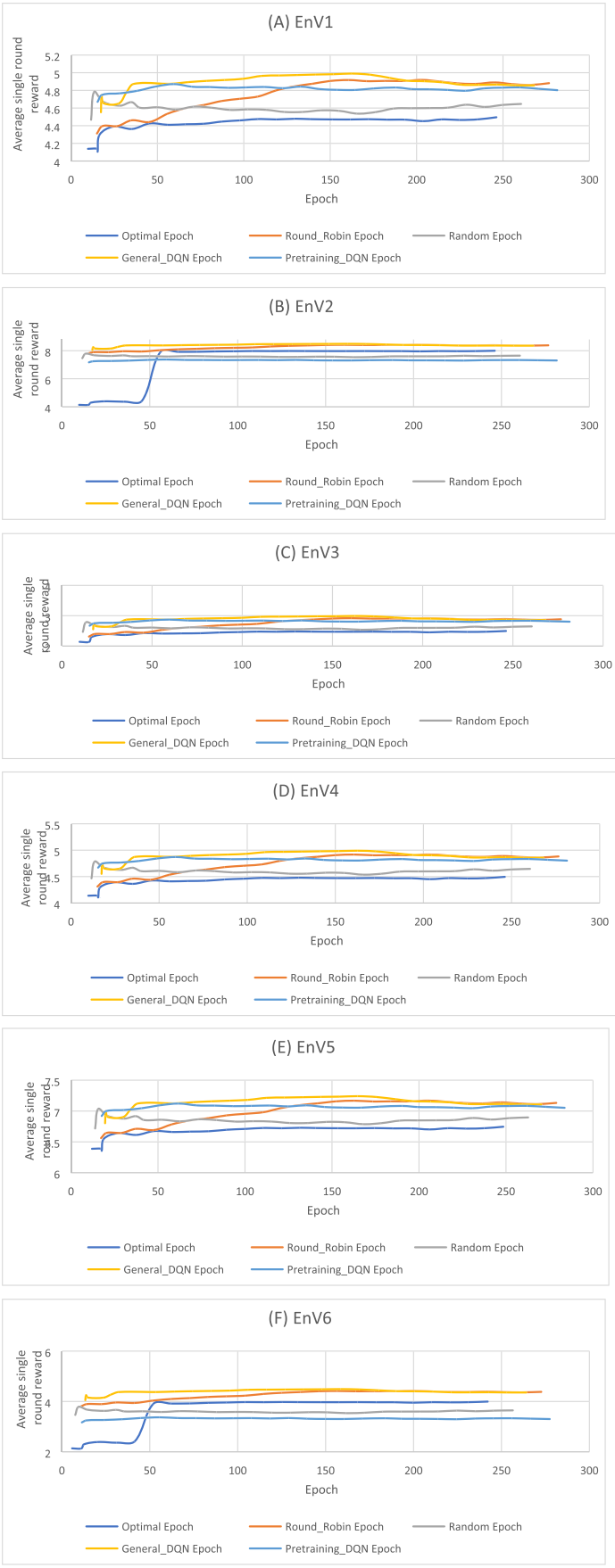

FIGURE 8 Comparison of scene-changing capabilities in various environments.

1. The throughput and working time per unit time during regular work, that is, the size of the reward function.
2. The performance at the initial stage of training, that is, the reward function's size at the initial training phase.
3. Changing scene learning ability, that is, to re-examine the above two indicators in a new environment.

5.2 | Simulation results

This section divides the experiment into three parts. The first part tests the strategy's performance, the second part tests the effect of the stretching system prelearning algorithm and the third part tests the learning ability of the process under changing scenarios.

In the first part of the experiments, the corresponding training strategies and other comparison strategies are tested in six test environment models. The experimental outcomes are illustrated in Figure 6A–F. The strategy that can be seen from the figure in this paper can obtain the highest reward function value. Although the optimal design can be close to our method in most cases, it is far below the results obtained by our way in some exceptional circumstances (as shown in Figure 6E). Value of the reward function, as the approach outlined in this paper closely approximates the optimal selection. Consequently, it allows for cautious transmission as energy levels approach depletion, preventing any undesired augmentation of points within the idle system. In this case, all algorithms have relatively comparable performance under poorly integrated channel states, as shown in Figure 6C. The comparison of comprehensive investigational results shows that the method in this paper can be used in the system to obtain better performance under the evaluation index of the difference between long-term average throughput and turnaround time.

In the second part of the experiment, the difference in the reward value of the six test environment models at the initial stage of training is compared with and without the pretraining algorithm. Figure 7 displays the outcomes of the experiments. As the figure shows, the suggested pretraining method is able to drastically cut down on the loss in the early stage of training and reach a better level in the early stage of training. In addition, movement, and in some environments, significantly improves convergence speed. Comparing comprehensive experimental results shows that the pretraining algorithm has a perfect effect.

In the third part of the experiment, the untrained and untrained policies using the pretrained algorithm are compared with the reward values of the traditional algorithm in six test environment models to test its learning ability in changing scenarios—experimental results, as shown in Figure 8.

In this paper, the investigation of scene change learning ability is divided into two steps. The first step is to check the model's performance in different environmental models; the second is to examine the acceleration effect of model pretraining in other ecological models. Figure 8 shows that our strategy can provide the best or close to the best performance after the environment model is changed and has a better learning rate boost. Furthermore, the comprehensive comparative experimental results show that the strategy in this paper has better results in different environmental models—good learning ability of scene change.

6 | CONCLUSION

In this article, we treat the power distribution issue in energy-harvesting wireless communication networks as a zero-information Markov decision process problem. Maximising both the total amount of time the system is operational, and its typical output over time is the objective. In order to determine the best approach to access management, this article applies DQN from the field of reinforcement learning. In addition, since the model of the environment is always evolving, this article suggests a time-invariant structural pretraining method to speed up the convergence of the access control strategy. The experimental findings demonstrate that the proposed pretraining technique can greatly decrease the loss in the initial stage of training and that the design outperforms the conventional strategy. As a result, the mechanism is highly adaptable to new conditions. Future work includes examining scalability to larger networks and exploring advanced machine learning techniques for improved performance and adaptability.

DATA AVAILABILITY STATEMENT
Research data are not shared.

ORCID

Mukesh Soni https://orcid.org/0000-0002-9228-6071
Renato R. Maaliw III https://orcid.org/0000-0002-7310-2708

REFERENCES

1. Song D, Shin W, Lee J. A maximum throughput design for wireless powered communication networks with IRS-NOMA. *IEEE Wirel Commun Lett*. 2021;10(4):849-853. doi:10.1109/LWC.2020.3046722
2. Ma W, Wang W, Jiang T. Joint energy harvest and information transfer for energy beamforming in backscatter multiuser networks. *IEEE Trans Commun*. 2021;69(2):1317-1328. doi:10.1109/TCOMM.2020.3036049
3. Özyurt S, Coşkun AF, Büyükçorak S, Karabulut Kurt G, Kucur O. A survey on multiuser SWIPT communications for 5G+. *IEEE Access*. 2022;10:109814-109849. doi:10.1109/ACCESS.2022.3212774
4. Zeng P, Wu Q, Qiao D. Energy minimization for IRS-aided WPCNs with non-linear energy harvesting model. *IEEE Wirel Commun Lett*. 2021;10(11):2592-2596. doi:10.1109/LWC.2021.3109642
5. Fang Z, Wang J, Ren Y, Han Z, Poor HV, Hanzo L. Age of information in energy harvesting aided massive multiple access networks. *IEEE J Select Areas Commun*. 2022;40(5):1441-1456. doi:10.1109/JSAC.2022.3143252
6. Su B, Ni Q, Yu W, Pervaiz H. Optimizing computation efficiency for NOMA-assisted mobile edge computing with user cooperation. *IEEE Trans Green Commun Network*. 2021;5(2):858-867. doi:10.1109/TGCN.2021.3056770
7. Li M, Zhou X, Qiu T, Zhao Q, Li K. Multi-relay assisted computation offloading for multi-access edge computing systems with energy harvesting. *IEEE Trans Vehic Technol*. 2021;70(10):10941-10956. doi:10.1109/TVT.2021.3108619
8. Cao K, Wang B, Ding H, et al. Improving physical layer security of uplink NOMA via energy harvesting jammers. *IEEE Trans Inf Forensics Secur*. 2021;16:786-799. doi:10.1109/TIFS.2020.3023277
9. Zhai C, Li Y, Wang X, Yu Z. Nonorthogonal multiple access with energy harvesting-based alternate relaying. *IEEE Syst J*. 2022;16(1):327-338. doi:10.1109/JSYST.2020.3034247
10. Khazali A, Tarchi D, Shayesteh MG, Kalbkhani H, Bozorgchenani A. Energy efficient uplink transmission in cooperative mm wave NOMA networks with wireless power transfer. *IEEE Trans Vehic Technol*. 2022;71(1):391-405. doi:10.1109/TVT.2021.3124076
11. Li B, Zhang M, Rong Y, Han Z. Transceiver optimization for wireless powered time-division duplex MU-MIMO systems: non-robust and robust designs. *IEEE Trans Wirel Commun*. 2022;21(6):4594-4607. doi:10.1109/TWC.2021.3131595
12. Mitsiou NA, Gavriilidis PN, Diamantoulakis PD, Karagiannidis GK. Wireless powered multiaccess edge computing with slotted ALOHA. *IEEE Commun Lett*. 2023;27(1):273-277. doi:10.1109/LCOMM.2022.3211190
13. Lei R, Xu D. On the outage performance of JT-CoMP-CNOMA networks with SWIPT. *IEEE Commun Lett*. 2021;25(2):432-436. doi:10.1109/LCOMM.2020.3029776
14. Singh CK, Singh V, Upadhyay PK, Lin M. Energy harvesting in overlay cognitive NOMA systems with hardware impairments. *IEEE Syst J*. 2022;16(2):2648-2659. doi:10.1109/JSYST.2021.3082552
15. Li N, Xiao M, Rasmussen LK, Hu X, Leung VCM. On resource allocation of cooperative multiple access strategy in energy-efficient industrial internet of things. *IEEE Trans Industr Inform*. 2021;17(2):1069-1078. doi:10.1109/TII.2020.2988643
16. Pei X, Duan W, Wen M, Wu Y-C, Yu H, Monteiro V. Socially aware joint resource allocation and computation offloading in NOMA-aided energy-harvesting massive IoT. *IEEE Internet Things J*. 2021;8(7):5240-5249. doi:10.1109/JIOT.2020.3034380
17. Shukla AK, Singh V, Upadhyay PK, Kumar A, Moualeu JM. Performance analysis of energy harvesting-assisted overlay cognitive NOMA systems with incremental relaying. *IEEE Open J Commun Soc*. 2021;2:1558-1576. doi:10.1109/OJCOMS.2021.3093671
18. Zhang J, Xie G, Han G, Yu ZL, Gu Z, Li Y. Compressive sensing-based power allocation optimization for energy harvesting IoT nodes. *IEEE Trans Wirel Commun*. 2022;21(6):4535-4548. doi:10.1109/TWC.2021.3131159
19. Chu Z, Zhong J, Xiao P, et al. RIS assisted wireless powered IoT networks with phase shift error and transceiver hardware impairment. *IEEE Trans Commun*. 2022;70(7):4910-4924. doi:10.1109/TCOMM.2022.3175833
20. Ting X, Zhao M, Yao X, Zhu Y. An improved communication resource allocation strategy for wireless networks based on deep reinforcement learning. *Comput Commun*. 2022;188:90-98, ISSN 0140-3664. doi:10.1016/j.comcom.2022.02.018
21. Wang Y, Shang F, Lei J, Zhu X, Qin H, Wen J. Dual-attention assisted deep reinforcement learning algorithm for energy-efficient resource allocation in industrial internet of things. *Future Gener Comput Syst*. 2023;142:150-164, ISSN 0167-739X. doi:10.1016/j.future.2022.12.009
22. Wang Y, Shang F, Lei J. Multi-granularity fusion resource allocation algorithm based on dual-attention deep reinforcement learning and lifelong learning architecture in heterogeneous IIoT. *Inf Fusion*. 2023;99:101871, ISSN 1566-2535. doi:10.1016/j.inffus.2023.101871
23. Wang Y, Shang F, Lei J. Energy-efficient and delay-guaranteed routing algorithm for software-defined wireless sensor networks: a cooperative deep reinforcement learning approach. *J Netw Comput Appl*. 2023;217:103674, ISSN 1084-8045. doi:10.1016/j.jnca.2023.103674
24. Yu F, Yang D, Wu F, Wang Y, He H. Resource optimization for UAV-assisted mobile edge computing system based on deep reinforcement learning. *Phys Commun*. 2023;59:102107, ISSN 1874-4907. doi:10.1016/j.phycom.2023.102107
25. Wang X, Zhang Y, Wu H, Liu T, Xu Y. Deep transfer reinforcement learning for resource allocation in hybrid multiple access systems. *Phys Commun*. 2022;55:101923, ISSN 1874-4907. doi:10.1016/j.phycom.2022.101923
26. Zhao J, Fajin H, Li J, Nie Y. Multi-agent deep reinforcement learning based resource management in heterogeneous V2X networks. *Digital Commun Netw*. 2023, ISSN 2352-8648. doi:10.1016/j.dcan.2023.06.003

27. Jain V, Kumar B, Gupta A. Cybertwin-driven resource allocation using deep reinforcement learning in 6G-enabled edge environment. *J King Saud Univ - Comput Inf Sci*. 2022;34(8, Part B):5708-5720, ISSN 1319-1578. doi:10.1016/j.jksuci.2022.02.005
28. Aghapour Z, Sharifian S, Taheri H. Task offloading and resource allocation algorithm based on deep reinforcement learning for distributed AI execution tasks in IoT edge computing environments. *Comput Netw*. 2023;223:109577, ISSN 1389-1286. doi:10.1016/j.comnet.2023.109577
29. Dutta H, Bhuyan AK. Subir Biswas, reinforcement learning based flow and energy management in resource-constrained wireless networks. *Comput Commun*. 2023;202:73-86, ISSN 0140-3664. doi:10.1016/j.comcom.2023.02.011
30. Supraja G, Veeranan J. Throughput maximization and reliable wireless communication in NOMA using chained fog structure and weighted energy efficiency power allocation approach. *Comput Commun*. 2023;208:147-157, ISSN 0140-3664. doi:10.1016/j.comcom.2023.05.024
31. Basit A, Wakeel A, Ahmad A, et al. Optimum power allocation for an energy harvesting wireless communication system considering energy storage losses. *Ad Hoc Netw*. 2023;144:103138, ISSN 1570-8705. doi:10.1016/j.adhoc.2023.103138
32. Huang J, Yang F, Chakraborty C, et al. Opportunistic capacity based resource allocation for 6G wireless systems with network slicing. *Future Gener Comput Syst*. 2023;140:390-401, ISSN 0167-739X. doi:10.1016/j.future.2022.10.032
33. Shi KL, Wang F. Joint task processing/offloading mode selection and resource-allocation for backscatter-aided and wireless-powered MEC. *Comput Netw*. 2023;224:109584, ISSN 1389-1286. doi:10.1016/j.comnet.2023.109584
34. Hu B, Chen L, Chen S. Joint trajectory-resource optimization for UAV-enabled uplink communication networks with wireless backhaul. *Comput Netw*. 2023;229:109779, ISSN 1389-1286. doi:10.1016/j.comnet.2023.109779

How to cite this article: Shukla K, Kollu A, Panwar P, et al. Energy-efficient resource allocation over wireless communication systems through deep reinforcement learning. *Int J Commun Syst*. 2023;e5589. doi:10.1002/dac.5589

FACULTY POSITION RECLASSIFICATION FOR SUCS

(DBM-CHED Joint Circular No. 3, series of 2022)

CERTIFICATION OF PERCENTAGE CONTRIBUTION

(Research Output with Multiple Authors)

Title of Research: Energy-efficient Resource Allocation over Wireless Communication Systems through Dee Reinforcement Learning

Type of Research Output: Journal Article (Scopus-Indexed, Wiley Publication)

Instruction: Supply ALL the names of the authors involved in the publication of Research output and indicate the contribution of each author in percentage. Each author shall sign the conforme column if he/she agrees with the distribution. The conforme should be signed by all the authors in order to be considered. Please prepare separate Certification for each output.

	Name of Authors	Current Affiliations	% Contribution	Conforme *(Sign if you agree with the % distribution)*
1	Kirti Shukla	Galgotias University	12.50%	
2	Archana Kollu	PC College of Engineering & Research	12.50%	
3	Poonam Panwar	Markandeshwar University	12.50%	
4	Mukesh Soni	Chandigarh University	12.50%	
5	Latika Jindal	Medicaps University	12.50%	
6	Hemlata Patel	Medicaps University	12.50%	
7	Ismail Keshta	AlMaarefa University	12.50%	
8	Renato Maaliw III	SLSU	12.50%	
	** Should have a total of 100%*		100.00%	

Prepared by:

Renato R. Maaliw III, DIT
(Name and Signature)
Faculty

Certified by:

Nicanor L. Guinto, Ph.D
(Name and Signature)
Director, Office of Research Services

Preview (IJCS-23-0362)

From: msobaidat@gmail.com

To: soni.mukesh15@gmail.com, kirti.shukla@galgotiasuniversity.edu.in, kadarchna@gmail.com, rana.poonam1@gmail.com, latika.mehrotra@medicaps.ac.in, hemlata.patel@medicaps.ac.in, imohamed@mcst.edu.sa, rmaaliw@slsu.edu.ph

CC:

Subject: IJCS-23-0362 - Decision

Body: 16-Apr-2023

Dear Dr. Soni,

We recognise that the impact of the COVID-19 pandemic may affect your ability to return your revised manuscript to us within the requested timeframe. If this is the case, please let us know.

Manuscript ID IJCS-23-0362 entitled "Energy-Efficient Resource Allocation over Wireless Communication Systems through Deep Reinforcement Learning" which you submitted to International Journal of Communication Systems has been reviewed. The comments of the referee(s) are included at the bottom of this letter.

A revised version of your manuscript that takes into account the comments of the referee(s) will be reconsidered for publication.

Please note that submitting a revision of your manuscript does not guarantee eventual acceptance, and that your revision may be subject to re-review by the referee(s) before a decision is rendered.

You can upload your revised manuscript and submit it through your Author Center. Log into submission.wiley.com/journal/dac and click on My Submissions. Sort by journal and submission status to locate this manuscript, then click the "Revise submission" button to submit your revision.

When submitting your revised manuscript, you will be able to respond to the reviewer comments when asked to "Upload your Author Response".

All supplementary and additional files will be carried over when you submit a revised manuscript. You may be required to provide additional files at the revision stage. If indicated to do so, please upload any additional required files as needed.

- Please include a Graphical Abstract: Authors must submit an abstract figure (diagram or illustration selected from the manuscript or an additional "eye-catching" figure) and accompanying text for this abstract with the original manuscript. The GTOC entry should include the paper title, the authors' names (with the corresponding author indicated by an asterisk) together with the figure and no more than 80 words or 3 sentences of text summarizing the key findings presented in the paper.

>> Our Expects Data Policy requires a Data Availability Statement, even if no data are available, please provide a statement in the ScholarOne.
https://authorservices.wiley.com/author-resources/Journal-Authors/open-access/data-sharing-citation/data-sharing-policy.html
- Please include the same data availability statement in the manuscript as the last reference in the list or before the reference section.
Please note that this statement will be published alongside your manuscript, if it is accepted for publication. https://authorservices.wiley.com/author-resources/Journal-Authors/open-access/data-sharing-citation/data-sharing-policy.html#standardtemplates

(If the above link space is blank, it is because you submitted your original manuscript through our old submission site. Therefore, to return your revision, please go to our new submission site here (submission.wiley.com/journal/dac) and submit your revision as a new manuscript; answer yes to the question "Are you returning a revision for a manuscript originally submitted to our former submission site (ScholarOne Manuscripts)? If you indicate yes, please enter your original manuscript's Manuscript ID number in the space below" and including your original submission's Manuscript ID number (IJCS-23-0362) where indicated. This will help us to link your revision to your original submission.)

Wiley Editing Services Available to All Authors
Should you be interested, Wiley Editing Services offers expert help with manuscript, language, and format editing, along with other article preparation services. You can learn more about this service option at www.wileyauthors.com/eeo/preparation. You can also check out Wiley's collection of free article preparation resources for general guidance about writing and preparing your manuscript at www.wileyauthors.com/eeo/prepresources.

This journal offers a number of license options, information about this is available here: https://authorservices.wiley.com/author-resources/Journal-Authors/licensing/index.html. All co-authors are required to confirm that they have the necessary rights to grant in the submission, including in light of each co-author's funder policies. For example, if you or one of your co-authors received funding from a member of Coalition S, you may need to check which licenses you are able to sign.

Once again, thank you for submitting your manuscript to International Journal of Communication Systems and I look forward to receiving your revision.

Sincerely,

Prof. Mohammad S. Obaidat
Editor-in-Chief, International Journal of Communication Systems

Referee(s)' Comments to Author:

Reviewer: 1

Comments to the Author
The paper seems to be very timely and focuses on a topic that is very paramount as it relates to energy efficient resource allocation over wireless communication systems through deep reinforcement learning and other deep learning architectures. The topic presented is good, but this paper still suffers from several issues as given below:

1. In the Introduction section of the research article, the authors must describe some research gaps as it relates with Energy-Efficient Resource Allocation over Wireless Communication Systems through Deep Reinforcement Learning.

2. The authors should improve on the literature review section of this manuscript. In section I, a table should be developed which compares the contributions of this work to other recent manuscripts in this field especially review papers where the use of deep reinforcement learning in wireless communication systems has been described and discussed. The focus and coverage of the work, its limitations should also be included in the table.

3. In section 2.1 Channel model, author needs to define symbol used in equation (1).

4. In section 3.2.2 Deep Q-learning, author needs to define significance of $Q(s, a, \theta) \approx Q(s, a)$ used in equation (6)

5. In 4.1 Simulation experiment setup, Author needs to mention the hyper parameter reward function value used in experiment process.

6. Test environment settings, Author needs to mention the Poisson distribution mean (λ) value used in experimental process.

7. In the conclusion section, a brief summary of the key findings from the research is requested. It is imperative to state the key takeaways from the work. In my opinion, the current conclusion is limited. What do you think?

8. So many abbreviations, the author should create a list of abbreviations of all abbreviated terms. This will help the flow and the readability of the manuscript.

9. More authoritative references on this subject should be cited. The references need to be increased.

Reviewer: 2

Comments to the Author
Comment 1. The weight of the manuscripts will be increased by placing a contribution on the research paper.
Comment 2. In section 2.1, the authors must describe the X_qand X_k parameter used in the proposed model
Comment 3. In section 3 , Author used SoftMax classifier in equation 20 , mention the X and c .
Comment 4. In section 4.2,Evaluation index and parameter setting author mention the hyper parameter in table 3 , author also specify the size of convolution kernel filter and dropout value.
Comment 5. In conclusion section mention the future direction for the proposed paper

Reviewer: 3

Comments to the Author
Comment 1. In the Introduction section of research article, the authors need to discus objective for Energy-Efficient Resource Allocation over Wireless Communication Systems.
Comment 2. Authors Need to gives an bird eye on advantage of Deep Reinforcement Learning for designing Energy-Efficient Resource Allocation over Wireless Communication Systems through Deep Reinforcement Learning.
Comment 3 The structure of an abstract typically includes a background, objectives, methods, results, conclusions, and keywords, which provide a concise summary of the research study, at hour need to update as above.
Comment 4 : Author need to discuss Multi-channel wireless communication model with energy harvesting procedure in brif in section 2
Comment 5: The paper could benefit from more detailed information on the methods and results of the quality improvement initiative.
Comment 6: It would be helpful to include more information on the specific interventions used, the data collection and analysis methods employed, and the results of the initiative in terms of improved outcomes or processes.
Comment 7: Additionally, the paper could benefit from a more in-depth discussion of the limitations and challenges faced during the quality improvement initiative, as well as potential future directions for improvement efforts.

Editor's Comments to the Author:

Associate Editor
Comments to the Author:
(There are no comments.)

Date Sent: 16-Apr-2023

Close Window