Hindawi
Mobile Information Systems
Volume 2021, Article ID 1430512, 9 pages
https://doi.org/10.1155/2021/1430512

Research Article

Energy-Efficient UAV Trajectory Design with Information Freshness Constraint via Deep Reinforcement Learning

Xinmin Li,[1] **Jiahui Li**,[1] **and Dandan Liu**[2]

[1]*School of Information Engineering, Southwest University of Science and Technology, Mianyang 621000, China*
[2]*College of Weapon Engineering, Naval University of Engineering, Wuhan 430030, China*

Correspondence should be addressed to Dandan Liu; liudandan_nue@126.com

Received 27 October 2021; Accepted 30 November 2021; Published 23 December 2021

Academic Editor: Han Wang

Unmanned aerial vehicle (UAV) technique with flexible deployment has enabled the development of Internet of Things (IoT) applications. However, it is difficult to guarantee the freshness of information delivery for the energy-limited UAV. Thus, we study the trajectory design in the multiple-UAV communication system, in which the massive ground devices send the individual information to mobile UAV base stations under the demand of information freshness. First, an energy-efficiency (EE) maximization optimization problem is formulated under the rest energy, safety distance, and age of information (AoI) constraints. However, it is difficult to solve the optimization problem due to the nonconvex objective function and unknown dynamic environment. Second, a trajectory design based on the deep Q-network method is proposed, in which the state space considering energy efficiency, rest energy, and AoI and the efficient reward function related with EE performance are constructed, respectively. Furthermore, to avoid the dependency of training data for the neural network, the experience replay and random sampling for batch are adopted. Finally, we validate the system performance of the proposed scheme. Simulation results show that the proposed scheme can achieve a better EE performance compared with the benchmark scheme.

1. Introduction

With the explosive increasing of global mobile devices and connections in the future wireless network, Cisco forecasts that the global mobile data traffic will reach 77 exabytes per month by 2022 [1], which is almost two times over the data traffic in 2020. To meet the needs of high volume of data traffic and massive connections, the sixth generation (6G) wireless communication system enables some promising technologies to improve the communication rate, enhance the wide coverage, access the massive devices, and strengthen the intelligence and security [2, 3]. The unmanned aerial vehicle (UAV) communication working as one of the promising technologies, which has the advantages of flexible deployment, controllable maneuver, and low cost, becomes an interesting topic in the industry and academia to drive the development of Internet of Things (IoT) applications [4–7].

Due to the influence of the UAV's trajectory on the rate performance and energy consumption directly, how to design the UAV's trajectory is vital in the various types of communication scenarios. Although there exists some literatures investigating the trajectory design [8–10] for single-UAV communication systems under different settings, multiple-UAV may serve the specific area to provide better communication rate and coverage performance, which can increase the interference level of the UAV communication network. In [11], the authors analyzed the influence of UAVs' positions on the rate performance and obtained the optimal positions in the two-UAV interference channel. The authors in [12] designed the trajectory of the multiple-UAV based on the successive convex approximation (SCA) method considering the backhaul. However, exploring the energy-efficient trajectory design is vital for the energy-limited UAV to enhance the sustainable communication capability. In [13], the comprehensive energy consumption

model including the communication energy and propulsion energy was proposed for rotary-wing UAV. The SCA-based technique was employed to optimize the UAV trajectory. Furthermore, a joint of scheduling the backscatter devices, trajectory design, and transmit power was proposed in [14] to maximize energy-efficiency (EE) performance. In addition, the authors in [15] optimized a constructed trajectory to minimize the propulsion energy of fixed-wing UAV. Above literatures focus on the UAV communication to provide the high performance communication link and enable the information delivery.

With the increase in the real-time and computation-intensive applications, a new metric is required to satisfy the demand of information freshness beyond the scope of the delay time performance metric. To characterize the freshness for information delivery accurately, age of information (AoI) has been proposed in [16], which precisely describes the timeless of information updates from the original generation at the perspective of the receiver. In general, AoI is defined as the time gap between the observed time and the recent update time. Taking the generation time and information update into consideration, it is different from the conventional performance metrics, such as delay. In order to meet the demands of the various communications, AoI performance metric has been introduced to optimize the generation policies and user scheduling [17–20]. To guarantee the information freshness of the UAV communication system, the authors in [21] proposed a dynamic programming-based path planning to update the collected data to minimize the AoI value. Recently, some learning methods were proposed to make the decisions in various dynamic scenarios [22–24]. Under the UAV's energy constraint, the AoI optimization scheme based on reinforcement learning (RL) was proposed in [23] by optimizing the UAV's trajectory. In [24], the authors designed a joint trajectory and packet scheduling scheme based on deep reinforcement learning (DRL) approach to minimize the weighted AoI performance of the single-UAV system. However, how to design the trajectory that not only improves the energy efficiency but also guarantees the information update on time in the UAV communication system remains an open problem.

In this paper, we consider the UAVs' trajectory design in the multiple-UAV-enabled communication system to maximize energy-efficiency performance, in which each ground device sends the individual information to the corresponding UAV. First, we formulate an energy-efficient trajectory design optimization problem in multiple-UAV communication systems under the practical constraints, such as rest energy, safety distance, and AoI metric. However, it is difficult to solve this optimization problem due to the nonconvex objective function and the unknown spaces for UAV's trajectory. Second, a deep Q-network (DQN) method is proposed to optimize the UAV's trajectory and reduce the computational complexity. We design the state space considering the UAV's position, energy efficiency, rest energy, and AoI and construct the efficient reward function related with the objective function and the constraints. Furthermore, to avoid the dependency of training data for the neural network, the experience replay and random sampling for batch are adopted, which can grantee the stability of the proposed scheme in the dynamic environment. Finally, we verify the system performance of the proposed scheme. Simulation results show that the proposed scheme outperforms the benchmark scheme.

The rest of paper is organized as follows. In Section 2, the system model and energy-efficiency optimization problem in the multiple-UAV communication system are presented. In Section 3, we present the proposed DQN scheme to solve the optimization problem in detail. The simulation results of the proposed scheme are shown in Section 4. Finally, we conclude the work in Section 5.

2. System Model

We consider a UAV communication system consisting of M single-antenna UAV base stations denoted as $\mathcal{M} = \{1, \ldots, M\}$ and K single-antenna IoT device denoted as $\mathcal{K} = \{1, \ldots, K\}$ (e.g., smart grid, agricultural, safety, or geographic information). All the UVAs working as the aerial base stations serve the ground devices depicted in Figure 1. The UAVs service range is within a specific radius, and the flight height is fixed as H.

For simplicity, $\mathbf{q}_m^u(t) = (x_m^u(t), y_m^u(t), H)$ and $\mathbf{q}_k^d = (x_k^d(t), y_k^d(t), 0)$ denote the three-dimensional (3D) position of the m-th UAV at time t and the k-th device's position, respectively. Thus, the 3D distance between the k-th device and the m-th UAV is written as

$$\begin{aligned} d_{k,m}(t) &= \left\| \mathbf{q}_m^u(t) - \mathbf{q}_k^d \right\| \\ &= \sqrt{\left(x_m^u(t) - x_k^d\right)^2 + \left(y_m^u(t) - y_k^d\right)^2 + H^2}. \end{aligned} \tag{1}$$

In order to model the channel information of the UAV communication link more practically, we adopt the probabilistic line-of-sight (LoS) channel model proposed in [25]. The probability of LoS link for the communication link between the k-th device and the m-th UAV, which is related with the device's elevation, can be expressed as

$$P_{k,m}^{LoS}\left(\alpha_{k,m}(t)\right) = \frac{1}{1 + a_1 \exp\left(-a_2\left[\alpha_{k,m}(t) - ta_2\right]\right)}, \tag{2}$$

where a_1 and a_2 are the channel parameters related with the environment of the communication link and $\alpha_{k,m}(t)$ denotes the elevation angle between the k-th device and the m-th UAV at time t. Thus, the probability $P_{k,m}^{NLoS}$ for non-line-of-sight (NLoS) link between the k-th device and the m-th UAV is $P_{k,m}^{NLoS}(\alpha_{n,i}(t)) = 1 - P_{k,m}^{LoS}(\alpha_{k,m}(t))$. Therefore, the average channel power gain between the k-th devices and the m-th UAV is defined as follows [25]:

$$h_{k,m} = P_{k,m}^{LoS}\left(\alpha_{k,m}\right) PL_{k,m}^{LoS} + P_{k,m}^{NLoS}\left(\alpha_{k,m}\right) PL_{k,m}^{NLoS}. \tag{3}$$

In this work, we adopt the time-division multiple access to serve the ground devices, in which the interference is originated from the devices using the same time resource. Thus, the signal-to-interference-plus-noise-ratio of the k-th user at the m-th UAV at the time t is expressed as

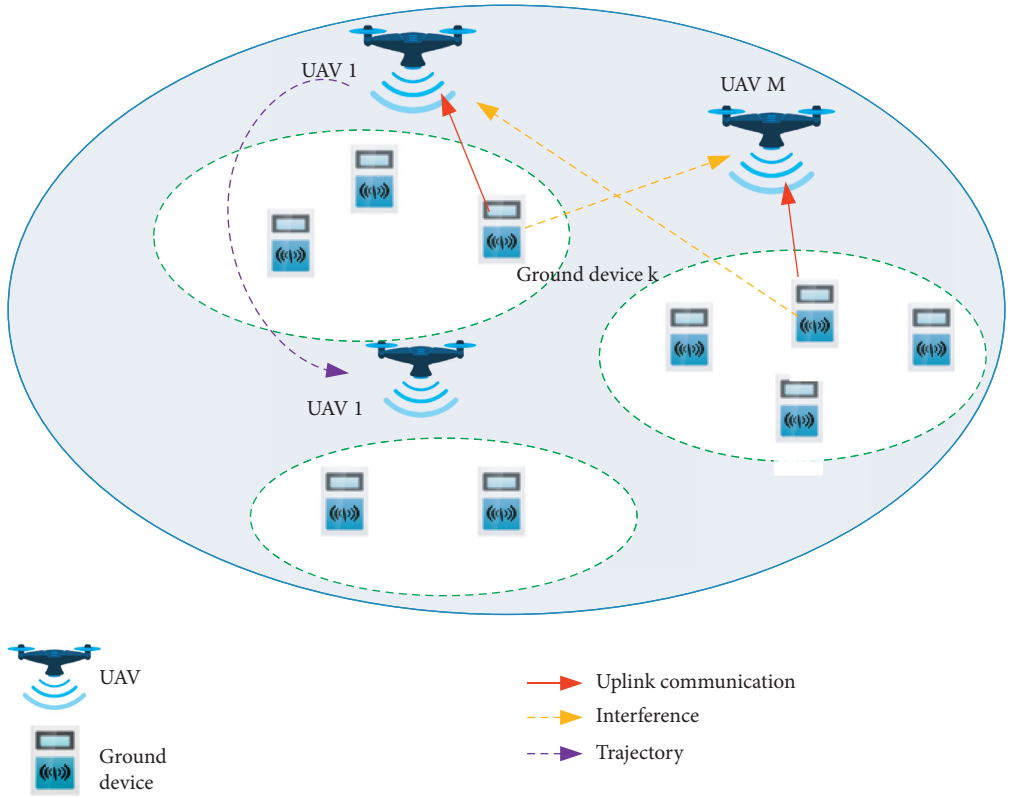

Figure 1: System model for UAV communication network serving massive IoT ground devices.

$$\gamma_{k,m}^{t} = \frac{p_k h_{k,m}}{\sum_{i \neq k} p_i h_{i,m} + \sigma^2}, \tag{4}$$

where p_k and σ^2 are the transmit power of the k-th ground device and the received noise power at each UAV, respectively. According to (4), the communication rate between the k-th device and the m-th UAV is written as

$$R_{k,m}^{t} = B \log_2\left(1 + \gamma_{k,m}^{t}\right), \tag{5}$$

where B denotes the transmission bandwidth. Since the UAV working as the aerial BS serves IoT devices with small data packets, UAV only requires comparably short time to receive the data, and we neglect the hover energy consumption. Thus, the energy consumption of UAV communication mainly consists of the communication energy consumption E_{com} and the propulsion energy consumption for flight E_{fly}. We assume that each UAV has a constant receiving power p^u, and the energy consumption consumed by the communication link from starting time to the time t is expressed as

$$E_{\text{com}}(t) = p^u t. \tag{6}$$

The energy consumption to support the UAV flight is the propulsion consumption. In general, the required power consumption can be modelled as follows [13]:

$$p_m^{\text{fly}}(t) = \left(c_1 \left\| v_m(t) \right\|^3 + \frac{c_2}{v_m(t)}\left(1 + \frac{\beta_m(t)}{g^2}\right)\right), \tag{7}$$

where $v_m(t)$ and $\beta_m(t)$ denote the flight speed of the m-th UAV and the flight acceleration of the UAV, respectively. The parameters c_1 and c_2 depend on the weight, the wing length of the UAV, and the air density in the flight environment, and g is the acceleration of gravity. Thus, the propulsion energy is $E_{m,\text{fly}}(t) = \int_0^t p_m^{\text{fly}}(\tilde{t})\mathrm{d}\tilde{t}$, and the energy consumption at the time t is written as

$$E_{m,\text{cons}}(t) = E_{\text{com}}(t) + E_{m,\text{fly}}(t). \tag{8}$$

Since the UAV has the limited energy for flight and communication, it is necessary to remain the enough rest energy for safe flight and return.

Let $E_{m,\max}$ denote the maximal energy of the m-th UAV; thus, the rest energy of the m-th UAV is defined as

$$E_m^r(t) = E_{m,\max} - E_{m,\text{cons}}(t). \tag{9}$$

In order to meet the essential safety for UAV flight, the rest energy E_m^r of UAV should not be less than the minimum rest energy $E_m^{r,th} = \chi E_{\max}$ with $\chi \in (0, 1)$ denoting the coefficient of rest energy.

The energy efficiency for UAV m to serve IoT devices at the time t is expressed as

$$\eta_m(t) = \frac{\sum_{k=1}^{K} R_{km}(t)}{p_m^{\text{fly}}(t) + p^u}. \tag{10}$$

AoI describes the age of the received packet at the destination to characterize the freshness of information collected by UAVs and becomes a new metric for the future

communication system [16]. For example, the device k sends different packets at the times $\tau_1, \tau_2, \ldots, \tau_n$ and the UAV m receives these packets at the times $\widetilde{\tau}_1, \widetilde{\tau}_2, \ldots, \widetilde{\tau}_n$. At time t, the immediate vicinity time for receiving the packets between UAV m and the device k is $\tau'_{km} = \max\{\widetilde{\tau}_1, \widetilde{\tau}_2, \ldots, \widetilde{\tau}_n\}$. Thus, AoI of UAV m is defined as the gap between the observed time and the maximal received time, which is written as

$$\Delta^t_{km} = t - \tau'_{km}. \tag{11}$$

It is noted that the smaller AoI, the higher freshness of the information while the bigger AoI, the less freshness of the information. For simplicity, $\Delta^0_{km} = 0, \forall k, m$, i.e., the AoI of the system at the time $t = 0$ sets zero. The average AoI is long-term metric to measure the total freshness of information over the duration T, which is expressed as

$$\overline{\Delta}_m = \frac{1}{T}\sum_{t=1}^{T}\sum_{k=1}^{K}\Delta^t_{km}. \tag{12}$$

The target is to maximize the energy efficiency of the UAV system by optimizing the trajectory under the AoI and energy constraints. Therefore, the optimization problem is expressed as follows:

$$\begin{aligned}(P1):\ &\max_{\mathbf{q}^u_m(t)}\sum_{m=1}^{M}\eta_m\\ &\text{s.t. } \left\|\mathbf{q}^u_m(t) - \mathbf{q}^u_{\overline{m}}(t)\right\| \geq d^{\text{th}}, \quad \overline{m} \in \mathscr{M}\backslash m,\end{aligned} \tag{13}$$

$$E^r_m \geq E^{r,\text{th}}, \quad m \in \mathscr{M}, \tag{14}$$

$$\overline{\Delta}_m \leq \overline{\Delta}^{\text{th}}, \quad m \in \mathscr{M}. \tag{15}$$

Constraints (13) and (14) originate from the safety requirements that UAVs' distance and rest energy all should have the minimum thresholds d^{th} and $E^{r,\text{th}}$ to avoid the flight conflict, while constraint (15) comes from the demand of the data freshness with the threshold $\overline{\Delta}^{\text{th}}$. The objective function is nonconvex, and there exists the mass of the computational complexity to search the serve sequence of the devices under AoI constraints and design the UAV trajectory with unknown spaces in the system; thus, it is challenging to handle the problem (P1). To solve the nonconvex problem, a reinforcement learning method is adopted by combining the deep neural network and reinforcement learning to design the UAVs' trajectories intelligently.

3. Proposed Solution Based on Deep Reinforcement Learning

RL is one of the potential machine learning since it can make decisions by choosing the beneficial action from the action space based on its past experiences in dynamic environments, in which the agent interacts with the environments and updates the rewards on the current state [26]. There are three fundamental parts in each RL algorithm: state of environment, action of agent, and the reward from the environment.

3.1. State, Action, and Reward Function. In this work, UAVs are regarded as the agents to decide the trajectory to satisfy the requirements of energy and distance, and the state space of the environment for UAV trajectory design can be defined as a five-dimensional state space including the UAV's position, energy efficiency, rest energy of UAV, and AoI of current state, respectively, i.e., the state of the m-th UAV is expressed as follows:

$$\mathbf{s}_m = \left[x^u_m(t), y^u_m(t), E_r(t), \eta_m(t), \Delta_m(t)\right]^T. \tag{16}$$

In the initial state, each UAV equips with the maximal energy and does not build the communication link with any devices. Thus, the initial state $\mathbf{s}_m = [x^u_m(0), y^u_m(0), E_{\max}, 0, 0]^T, m \in \mathscr{M}$, where AoI in the first time is set as zero for all UAVs. The rest energy, energy efficiency, and AoI in each state can be updated according to (9)–(11), respectively. If the rest energy is smaller than the minimal threshold, UAV makes a decision to the initial position to guarantee the safety of UAV and stops the update of the state.

At any state of environment, the UAV can select the flying direction to serve the ground devices or reduce the interference in the UAV communication system. For simplicity, the action space is set as $\mathscr{A} = \{0, \ldots, 2i/L * \pi, \ldots, 2(L-1)/L * \pi\}$, with L uniform directions in $[0, 2\pi)$. If the $L = 4$, the setting is typical with four orthogonal directions {left, right, frontward, backward}. After selecting the action a_t in the action space $\mathscr{A}$, the UAV can transit the current state $\mathbf{s}_t$ to the next state $\mathbf{s}_{t+1}$, e.g., the next position of the m-th UAV in the flying time duration δ_t:

$$\begin{aligned}x^u_m(t+1) &= x^u_m(t) + v(t)\delta_t\cos(a_t),\\ y^u_m(t+1) &= y^u_m(t) + v(t)\delta_t\sin(a_t).\end{aligned} \tag{17}$$

Based on the new UAV position $\mathbf{q}^u_m(t+1) = (x^u_m(t+1), y^u_m(t+1), H)$, the communication rate is obtained while the energy efficiency and AoI values can also be updated to show the performance of the new state. Since the reward function is significant to obtain a optimal policy, the system adopts the energy efficiency to construct the reward function, i.e.,

$$r_m(s_t, a_t) = \begin{cases}\phi\eta_m(t), & \text{if (13)} - \text{(15)are satisfied},\\ 0, & \text{otherwise},\end{cases} \tag{18}$$

where $0 \leq \phi \leq 1$ denotes the normalized parameter for energy efficiency (since the value of energy efficiency is comparatively high, the normalized parameter can balance the reward between negative and positive reward; in general, it is almost the order of magnitude of 10^{-3}, which is based on the reward value in the specific communication system). It is observed that the UAV will obtain the positive reward when the action a_t satisfies the corresponding constraints, while it will obtain any reward or penalty if the UAV violates the constraints. Since the reward function is monotonically increasing with respect to (w.r.t) the objective function in P1, the UAV makes the decision toward the energy-efficiency maximization, and the system can obtain the optimal solution.

Since the size of the state space is nonlinear with the size of UAV's position, action space, rest energy, and AoI, it exists the intensive computation complexity to obtain the optimal action by exhaustive search exists, which is impractical for the energy-limited UAV system. Moreover, the Q-learning method based on Q-table requires the large memory to store the Q-table, and it cannot be suitable for this optimization problem.

3.2. Control Policy and Deep Reinforcement Learning Algorithm. Recently, the deep RL algorithm has become a promising technique to tackle the resource allocation and performance optimization in the wireless communication systems. However, the continuous action space in the UAV communication system needs to be quantized into discrete and formulate the state-action function to characterize the influence of the selected action on the performance with a specific state. The deep RL method is applied to solve the problem because it can use the neural network to learn the policy to reduce the high dimensionality of the state space instead of storing the value. The detailed design framework based on DRL in the UAV communication system is shown in Figure 2. The Q-function originated from the Q-learning is adopted to maximize the long-term cumulated reward. Given the control policy ξ for the m-th UAV, the Q-function is defined as

$$Q^{\xi}(\mathbf{s}_t, a_t) = E\left[r_m(\mathbf{s}_t, a_t) + \sum_{j=1}^{t-1} \gamma^{t-j} r(\mathbf{s}_j, a_j)\right], \tag{19}$$

where $\gamma \in [0, 1]$ is the discount factor. If the discount factor $\gamma = 0$, the Q-function is only related with the current reward, i.e., the selected action only maximizes the current reward without considering the future reward. Thus, the optimal action to maximize the objective function in P1 can be written as follows:

$$a_t^* = \arg\max_{a_j \in \mathscr{A}} Q^{\xi}(\mathbf{s}_t, a_j). \tag{20}$$

To obtain the optimal control policy ξ^*, the Q-function is updated based on the Bellman equation:

$$Q(\mathbf{s}_t, a_t) = Q(\mathbf{s}_t, a_t) + \nu\left(r(\mathbf{s}_t, a_t) + \gamma \max_{a_j \in \mathscr{A}} Q(\mathbf{s}_{t+1}, a_j) - Q(\mathbf{s}_t, a_t)\right), \tag{21}$$

where ν is the learning rate. According to (21), the UAV can update the Q-function and learn the control policy based on the stored Q-values by the selecting the action to maximize the reward. However, there exists one puzzle during the processing of the learning: how to select the action in the limited state-action values. At the starting of learning, the UAV only has the some partial Q-values and cannot choose an appropriate action. Thus, the UAV should explore the environment sufficiently to obtain Q-values of all state-action pairs. To tackle this issue, an ϵ-greedy strategy is applied to explore the environment with the probability ϵ, which is written as

$$a = \begin{cases} \text{random}(\mathscr{A}), & \text{with probability } \epsilon, \\ \arg\max_{a_j \in \mathscr{A}} Q^{\xi}(\mathbf{s}_t, a_j), & \text{with probability } 1 - \epsilon. \end{cases} \tag{22}$$

At each state, the UAV can take a random action with the probability ϵ to explore the environment. As the number of exploration increases, the probability can decrease to guarantee the system performance with the UAV selecting the optimal action. Since the unknown state space for the UAV's trajectory may lead to a large memory size and a slow convergence rate, thus deep neural network is an effective method to extract features from the existing data sets intelligently and reduce the computational complexity by predicting the output in parallel. According to the framework in Figure 2, the tuple consists of the state, action, reward, and next state working as the input of deep neural network to output Q-value as $Q(s_{t+1}, a_t|\theta_i)$ in the estimate and target neural networks, where θ_i denotes the parameters of the neural network during i-th training. The target neural network is the replica of the estimate neural network every N_{rep} steps to make the two neural networks as close as possible to guarantee the stability. Therefore, it is important to optimize the parameters of neural network θ_i based on the suitable loss function to obtain the optimal Q-function. The loss function based on the error is defined as follows:

$$\mathscr{L}(\boldsymbol{\theta}_{i+1}) = \left|r(\mathbf{s}_t, a_t) + \gamma \max_{a_j \in \mathscr{A}} Q(\mathbf{s}_{t+1}, a_j|\boldsymbol{\theta}_i) - Q(\mathbf{s}_t, a_t|\boldsymbol{\theta}_i)\right|^2. \tag{23}$$

Based on the loss function and the training data set, some optimizers can be used to obtain the optimal parameters of neural network, such as gradient descent algorithm and Adam algorithm.

The training data are vital for training the deep neural network. However, there exist the following challenges: first, the UAV communication system is time-varying and the objective function is related with the UAV's trajectory and AoI. How to obtain the sufficient number of training data in the dynamic environment is crucial for optimizing the neural network. Second, empirical evidence demonstrated that independent training data can enhance the stability and improve the convergence for the neural network. Thus, obtaining independent training data is another challenge to optimize the neural network. To address these aforementioned challenges, the experience replay and random sampling method are adopted. For the fixed experience replay memory with the size N_{mem}, the training data will be updated every N_{tr} steps to replace the history data, which can automatically fresh the memory to obtain the fresh training data. To avoid the dependency of the stored data to train the neural network, the random sampling method is used to form the batch by choosing the experience from the replay memory randomly, which can smooth the changes between the history data and the new observation. The proposed DQN-based trajectory design scheme for energy-efficiency maximization is shown in Algorithm 1.

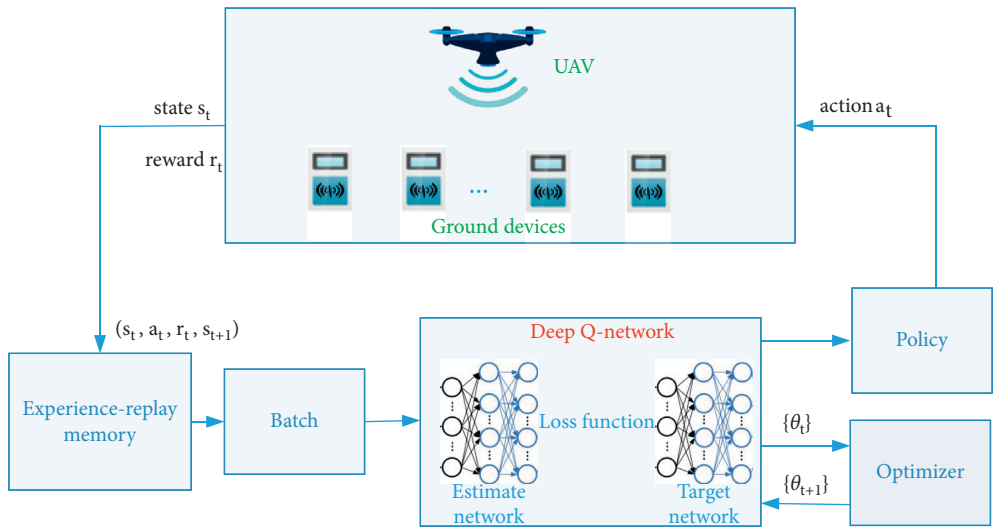

Figure 2: The design framework based on DRL for UAV's trajectory.

(1) Initialize learning parameters $\{\phi, \gamma, \epsilon\}$, memory size, batch size, the maximal episode $N_{\max}$, observed time T, and $N_{\text{step}} = 0$;
(2) **for** $n = 1$: $N_{\max}$ **do**
(3) Initialize the environment and state $\mathbf{s}_0$.
(4) **for** $t = 1$: T **do**
(5) If random() $> \epsilon$, select an action a_t using (20). Otherwise, select an random action from the action set $\mathcal{A}$.
(6) Execute the action a_t, compute EE, rest energy, and AoI, and obtain the next position $\mathbf{q}_m^u(t+1)$ to form the next state $\mathbf{s}_{t+1}$. According to (18), compute the reward. Store $\langle \mathbf{s}_t, a_t, r_t, \mathbf{s}_{t+1} \rangle$ into the experience-reply memory.
(7) If $N_{\text{step}} \% N_{\text{rep}} == 0$, duplicate the estimate neural network to target neural network.
(8) Train the neural network based on loss function in (23) to optimize the parameter θ, $N_{\text{step}} \leftarrow N_{\text{step}} + 1$.
(9) **end for**
(10) **end for**

Algorithm 1: Deep Q-network-based trajectory design scheme for energy-efficiency maximization.

4. Simulation Results

In this section, we consider three UAVs flying in the square area with $300 \times 300\,\text{m}^2$ to verify the proposed scheme. All ground devices are uniformly distributed in this area to transmit the independent information to UAVs during the observed time $T = 200$. For performance comparison, DQN without experience-reply is adopted as the benchmark scheme. Unless otherwise stated, the simulation parameters are shown in Table 1.

In order to train the deep neural network, three fully connected hidden layers with 100 hidden nodes are adopted. The size of experience-reply memory is 400 and is randomly selected from the memory to construct the batch. The gradient descent method is used to optimize the parameter of neural network. For the experience-reply memory, the new training data always update the oldest history data. To guarantee the size of efficient batch data, the training starts after 100 steps. Other parameters of the deep neural network are shown in Table 2. The simulation environment is Intel i7 CPU, Python 3.7, and TensorFlow 1.14 to train the UAV's deep neural network. All results are averaged over 500 episodes.

Figure 3 shows the effect of the flight speed v on the AoI performance. It is observed that AoI value decreases with increase in the flight speed, i.e., the information freshness can improve by increasing the UAV's flight speed, which comes from the fact that the frequency of the information update increases as the speed increases in the limited flying environment. Compared with the benchmark scheme, the AoI value decreases by 3.5% and 9.3% with $v = 20$ m/s and $v = 30$ m/s. However, since the propulsion power increases cubically w.r.t the flight speed, it is unwise only to increase the UAV's speed to maximize EE performance under the AoI constraint.

Figure 4 demonstrates the influence of the learning rate ν on the EE performance. We can find that the optimal EE performance of the proposed scheme can be achieved when the learning rate ν equals 0.4. Since the learning rate directly affects the convergence of the proposed scheme, the proposed scheme has a slow convergence rate to obtain the optimal EE performance. The proposed scheme can achieve

Table 1: Simulation parameters for the UAV communication system.

Symbol	Description	Value
B	Bandwidth	1 MHz
p	Transmit power of users	23 dBm
σ^2	Noise power	−100 dBm
K	Number of ground devices	30
M	Number of UAVs	3
L	Number of UAV's directions	10
H	UAV flight height	80 m
v	Flight speed	20 m/s
$E_{\max}$	Maximum energy of UAVs	1.5696×10^5 J
a_1, a_2	Channel parameters	[10, 0.6]
d^{th}	Distance threshold	1 m
χ	Coefficient of rest energy	0.1
Δ^{th}	AoI threshold	500

Table 2: The settings of the deep neural network.

Symbol	Description	Value
ν	Learning rate	0.4
ϵ	Greedy coefficient	0.05
γ	Discount factor	0.9
ϕ	Normalized parameter for EE	10^{-3}
N_{bat}	Batch size	64
N_{tr}	Interval of training data	3

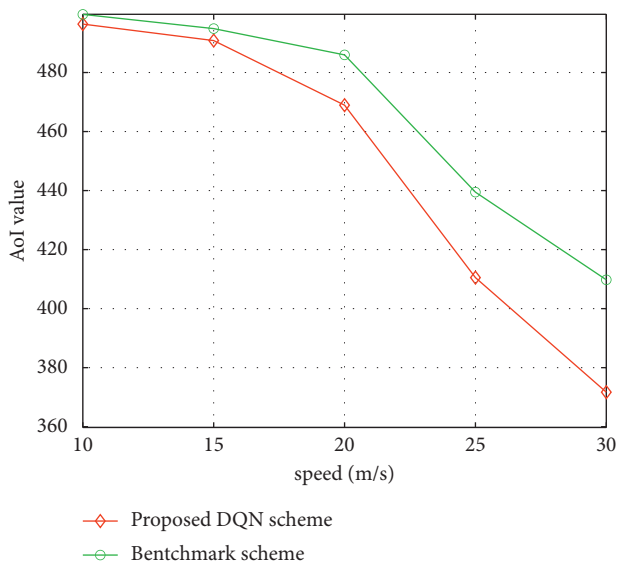

Figure 3: The effect of the flight speed on the AoI performance with $v = 0.4$.

22.2 Kbit/J at least compared with 21.7 Kbit/J when the learning rate becomes large.

Figure 5 shows the stability of the proposed scheme with the training number. It is noted that the loss fluctuates when the number of the training is small, and it converges to a small value as the number of training becomes large, which is result of the increasing number of the training batch. It is also found that the UAVs may have different convergence rates. There are two reasons as follows:

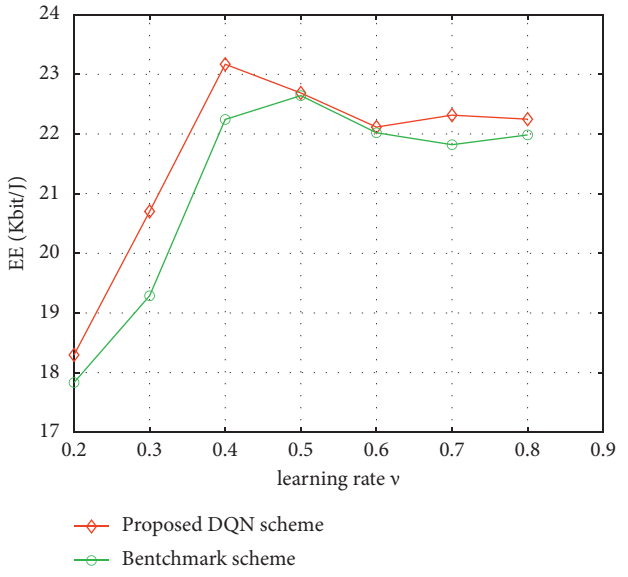

Figure 4: EE performance versus the learning rate with $v = 20$ m/s.

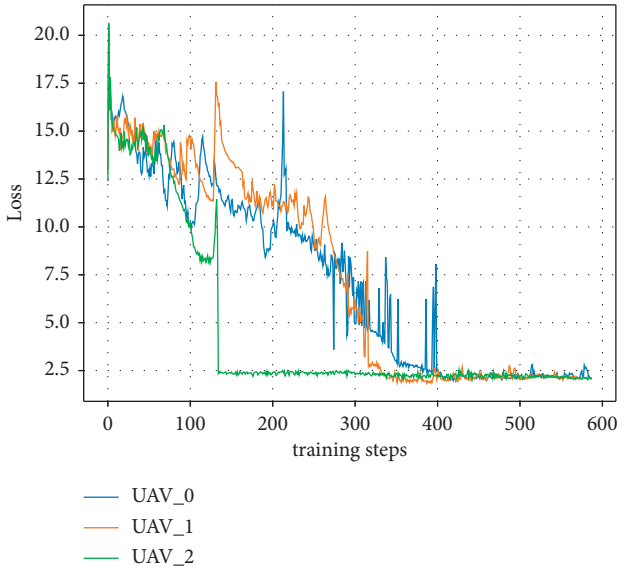

Figure 5: The effect of the training number on the loss function.

(1) UAVs have different initialized actions to result in the different estimate and target neural networks

(2) Each UAV has independent channel power gain related to the distance between UAV and served ground users to form the different state space and reward value

5. Conclusion

In this letter, we considered the UAV communication system to serve the IoT ground devices, which send the fresh information to the corresponding UAVs. Taking the rest energy and AoI into account, the DQN-based trajectory design was proposed to maximize the energy-efficiency performance. The state space related with the rest energy and AoI and the reward function related with EE performance are constructed, respectively. Under the experience replay and random sampling for batch, the simulation results show that the proposed DQN scheme achieves better performance compared with the benchmark scheme. In the future, considering the hover energy consumption for the massive IoT devices is a significant work while it is also interesting to design the collaborative UAV communication by exchanging the information for neural network optimization.

Data Availability

The data used to support the findings of this study are included within this article.

Conflicts of Interest

The authors declare there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported in part by the National Key R&D Program of China (2019YFB1705100), Foundation of Southwest University of Science and Technology (18zx7142 and 19xn0086), the Sichuan Science and Technology Program (2019JDTD0019), and China Scholarship Council Program (202008515123).

References

[1] Cisco, "Cisco visual networking index: Global Mobile Data traffic forecast update 2017–2022," Cisco, 2019, https://s3.amazonaws.com/media.mediapost.com/uploads/CiscoForecast.pdf.

[2] G. Flagship, *Key Drivers and Research Challenges for 6G Ubiquitous Wireless Intelligence*, G. Flagship, University of Oulu, Oulu, Finland, 2019, http://jultika.oulu.fi/files/isbn9789526223544.pdf.

[3] X. You, C. Wang, J. Huang et al., "Towards 6G wireless communication networks: vision, enabling technologies, and new paradigm shifts," *Science China*, vol. 64, pp. 1–74, 2021.

[4] S. Aggarwal, N. Kumar, and S. Tanwar, "Blockchain-envisioned UAV communication using 6G networks: open issues, use cases, and future directions," *IEEE Internet of Things Journal*, vol. 8, no. 7, pp. 5416–5441, 2021.

[5] J. Xu, Y. Zeng, and R. Zhang, "UAV-enabled wireless power transfer: trajectory design and energy optimization," *IEEE Transactions on Wireless Communications*, vol. 17, no. 8, pp. 5092–5106, 2018.

[6] H. Wang, X. Li, R. H. Jhaveri, and T. R. Gadekallu, "Sparse Bayesian earning ased hannel stimation in FBMC/OQAM industrial IoT networks," *Computer Communications*, vol. 176, pp. 40–45, 2021.

[7] Y. Huang, W. Mei, J. Xu, L. Qiu, and R. Zhang, "Cognitive UAV communication via joint maneuver and power control," *IEEE Transactions on Communications*, vol. 67, no. 11, pp. 7872–7888, 2019.

[8] J. Chen and D. Gesbert, "Optimal positioning of flying relays for wireless networks: a LOS map approach," in *Proceedings of the IEEE International Conference on Communications*, pp. 1–6, (ICC), Paris, France, May 2017.

[9] X. Li, Q. Li, D. Kong, X. Zhang, and X. Wang, "Learning based trajectory design for low-latency communication in UAV-enabled smart grid networks," in *Proceedings of the IEEE Vehicular Technology Conference*, pp. 1–5, VTC2020-Fall, Victoria, B.C., Canada, November 2020.

[10] Y. Zeng, Q. Wu, and R. Zhang, "Accessing from the sky: a tutorial on UAV communications for 5G and beyond," *Proceedings of the IEEE*, vol. 107, no. 12, pp. 2327–2375, 2019.

[11] X. Li and J. Xu, "Positioning optimization for sum-rate maximization in UAV-enabled interference channel," *IEEE Signal Processing Letters*, vol. 26, no. 10, pp. 1466–1470, 2019.

[12] P. Li and J. Xu, "Placement optimization for UAV-enabled wireless networks with multi-hop backhauls," *Journal of Communications and Information Networks*, vol. 3, no. 4, pp. 64–73, 2018.

[13] Y. Zeng, J. Xu, and R. Zhang, "Energy minimization for wireless communication with rotary-wing UAV," *IEEE Transactions on Wireless Communications*, vol. 18, no. 4, pp. 2329–2345, 2019.

[14] G. Yang, R. Dai, and Y.-C. Liang, "Energy-efficient UAV backscatter communication with joint trajectory design and resource optimization," *IEEE Transactions on Wireless Communications*, vol. 20, no. 2, pp. 926–941, 2021.

[15] F. Dong, L. Li, Z. Lu, Q. Pan, and W. Zheng, "Energy-efficiency for fixed-wing uav-enabled data collection and forwarding," in *Proceedings of the 2019 IEEE International Conference on Communications Workshops (ICC Workshops)*, pp. 1–6, Shanghai, China, May 2019.

[16] R. Y. S. Kaul and M. Gruteser, "Real-time status: how often should one update?" in *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM)*, pp. 2731–2735, Orlando, FL, USA, March 2012.

[17] Y. Sun, E. Uysal-Biyikoglu, R. D. Yates, C. E. Koksal, and N. B. Shroff, "Update or wait: how to keep your data fresh," *IEEE Transactions on Information Theory*, vol. 63, no. 11, pp. 7492–7508, 2017.

[18] I. Kadota, A. Sinha, and E. Modiano, "Optimizing age of information in wireless networks with throughput constraints," in *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM)*, pp. 1844–1852, Honolulu, HI,USA, April 2018.

[19] B. Zhou and W. Saad, "Joint status sampling and updating for minimizing age of information in the internet of things," *IEEE Transactions on Communications*, vol. 67, no. 11, pp. 7468–7482, 2019.

[20] R. Talak, S. Karaman, and E. Modiano, "Optimizing information freshness in wireless networks under general interference constraints," *IEEE/ACM Transactions on Networking*, vol. 28, no. 1, pp. 15–28, 2019.

[21] Z. Jia, X. Qin, Z. Wang, and B. Liu, "Age-based path planning and data acquisition in uav-assisted iot networks," in *Proceedings of the IEEE International Conference on Communications Workshops (ICC Workshops)*, pp. 1–6, Shanghai, China, May 2019.

[22] H. Wang, "Low-complexity MIMO-FBMC sparse channel parameter estimation for industrial big data communications," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 5, pp. 3422–3430, 2020.

[23] B. Yin, X. Li, X. Zhang, and L. Wei, "Trajectory optimization for age of information minimization in uav communication systems," in *Proceedings of the Accteped by IEEE Vehicular Technology Conference (VTC2021-Fall)*, Helsinki, Finland, April 2021.

[24] M. A. Abd-Elmagid, A. Ferdowsi, H. S. Dhillon, and W. Saad, "Deep reinforcement learning for minimizing age-of-information in uav-assisted networks," in *Proceedings of the IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, Waikoloa, HI, USA, December 2019.

[25] A. Al-Hourani, S. Kandeepan, and S. Lardner, "Optimal LAP altitude for maximum coverage," *IEEE Wireless Communications Letters*, vol. 3, no. 6, pp. 569–572, 2014.

[26] R. S. Sutton and A. G. Barto, *Reinforcement Earning: An Introduction*, MIT Press, Cambridge,MA, USA,, 2018.