한국정밀공학회지 제 42 권 제 3 호 pp. 197-202 March 2025 / 197
J. Korean Soc. Precis. Eng., Vol. 42, No. 3, pp. 197-202 http://doi.org/10.7736/JKSPE.024.099
ISSN 1225-9071 (Print) / 2287-8769 (Online)

UAV 통신 환경에서 심층 Q-네트워크 기반 무선 자원 할당 기법

Deep Q-network Based Resource Allocation for UAV Communication Systems

한동희[1,#], 김주영[1], 이상흥[1]
Dong Hee Han[1,#], Joo Young Kim[1], and Sang Heung Lee[1]

1 LIG넥스원㈜ (LIG Nex1 Co., Ltd.)
Corresponding Author / E-mail: donghee.han@lignex1.com, TEL: +82-031-8038-0246
ORCID: 0000-0002-9190-005X

KEYWORDS: Radio resource allocation (무선 자원 할당), Power allocation (전력 할당), Deep reinforcement learning (심층 강화 학습), UAV communications (무인기 통신)

In this paper, we propose a deep Q-network-based resource allocation method for efficient communication between a base station and multiple Unmanned Aerial Vehicles (UAVs) in environments with limited wireless resources. This method focused on maximizing the throughput of UAV to Infrastructure (U2I) links while ensuring that UAV to UAV (U2U) links could meet their data transmission time constraints, even when U2U links share the wireless resource used by U2I links. The deep Q-network agent uses the Channel State Information (CSI) of both U2U and U2I links, along with the remaining time for data transmission, as state, and determines optimal Resource Block (RB) and transmission power for each UAV. Simulation results demonstrated that the proposed method significantly outperformed both random allocation and CSI-based greedy algorithms in terms of U2I link throughput and the probability of meeting U2U link time constraints.

Manuscript received: August 20, 2024 / Revised: December 27, 2024 / Accepted: January 6, 2025
This paper was presented at KSPE Spring Conference in 2024

1. 서론

UAV (Unmanned Aerial Vehicle, 무인 항공기) 기반 통신 시스템은 효율적이고 유연한 네트워크 구축이 가능하여 최근 연구가 활발히 진행되고 있다. UAV 통신 시스템은 재난 지역의 긴급 물품 배송, 원격 의료 지원, 산불 감지, 감시 시스템 등 여러 분야에서 활용되고 있으며, 특히 통신 인프라가 제한적인 지역에서 효과적인 통신 수단을 제공할 수 있다는 장점이 있다[1-4]. UAV 통신 시스템은 주로 U2U (UAV-to-UAV, UAV 간 통신) 링크와 U2I (UAV-to-Infrastructure, 지상 인프라 간의 통신) 링크로 구성된다. U2U와 U2I 링크는 각기 다른 목적을 위해 사용되지만, 동일한 주파수 대역을 공유할 때 자원 간섭 문제와 신호 충돌이 발생할 수 있어 효율적인 자원 할당이 필수적이다.

기존 연구들은 다수의 지상 사용자와 다수의 UAV 간의 간섭을 최소화하면서 지상 사용자와 UAV 링크의 통신 품질 균형을 유지하는 방법[5], 다수의 지상 사용자와 다수의 UAV 간 통신 성능 개선을 위해 UAV의 경로 최적화 및 자원 할당 문제를 기계학습으로 개선하는 방법[6], 재해 지역에서의 채널 모델을 구성하고, 제안하는 채널 모델에 UAV의 통신 자원을 할당하는 방법[7] 등에 중점을 두고 있다.

본 연구는 기존 연구들과 달리 일반적인 접근을 통해, 단일 기지국과 U2I 링크 간 사용하는 주파수 대역을 U2U 링크가 공유하는 환경에서 U2I 링크의 Throughput을 최대화하고, U2U 링크의 실시간 데이터 전송을 보장하기 위해 심층 Q-네트워크 기반의 무선 자원 할당 기법을 제안한다. 심층 Q-네트워크 기반 접근은 다양한 환경에서 UAV의 자율적 의사결정을 가능하게 하여, 상호 간의 자원 간섭을 최소화하고 네트워크의 안정성을 높일 수 있다.

본 연구는 UAV 기반 통신 시스템이 높은 Throughput과 낮은 데이터 전송 지연을 통해 재난 대응, 물류 지원, 원격 감시 등의

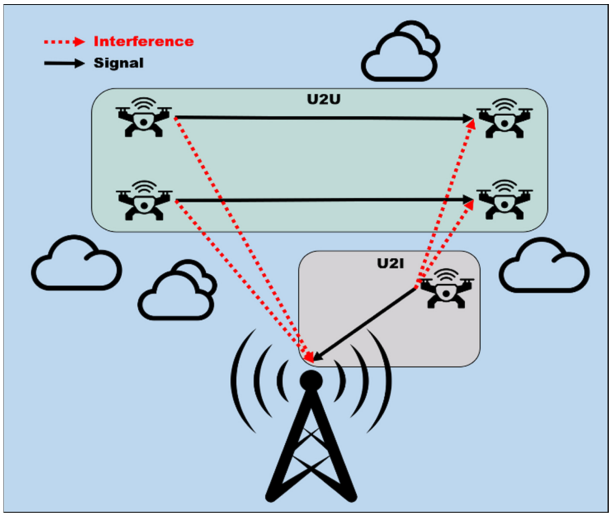

Fig. 1 System model

분야에서 향상된 통신 성능을 제공할 수 있다는 점에서 의의를 가진다. 특히, 심층 Q-네트워크 에이전트의 자율적인 자원 할당을 통해 기지국 의존도를 낮추고 네트워크 성능을 최대함으로써, 보다 다양한 상황에서 UAV의 실용성을 증진시킬 수 있다.

2. 시스템 모델

Fig. 1은 본 연구에서 가정한 단일 기지국과 다수의 UAV가 있는 환경을 나타낸다. 가정한 환경에서 통신 구조는 U2U 및 U2I 링크로 구성되어 있다. 그리고 U2U 링크는 U2I 링크에서 사용하는 RB (Resource Block, 무선 자원 블록)를 재사용한다. 여기서 기지국에 의해 형성된 U2I 링크의 수를 M, U2U 링크의 수를 K로 정의한다. 이와 같이 정의했을 때 m 번째 U2I 링크의 SINR (Signal-to-Interference-plus-Noise-Ratio, 신호 대 간섭·잡음비)과 Throughput 수식은 식(1)과 식(2)와 같다.

$$SINR^{I}[m] = \frac{P^{I}_{m} h_{m}}{\sigma^2 + \Sigma_{k \in K} p_k[m] p^{u}_{k} \tilde{h}_{k}} \tag{1}$$

$$C^{I}[m] = W \cdot log(1 + SINR^{I}[m]) \tag{2}$$

여기서 P^{I}_{m}는 m 번째 U2I 링크의 송신부에서 전송하는 전력이다. h_m은 m 번째 U2I 링크의 채널 이득이다. σ^2 은 잡음 전력이다. $p_k[m]$는 k 번째 U2U 링크의 송신부에서 m 번째 RB를 사용하는지 여부이며 0 또는 1의 값을 가진다. p^{u}_{k}는 k 번째 U2U 링크의 송신부에서 전송하는 전력이다. $\tilde{h}_k$는 k 번째 U2U 링크의 채널이득이다. $C^{I}[m]$는 m 번째 U2U 링크의 Throughput이다. W는 부반송파 대역폭이다. $SINR^{I}[m]$은 m 번째 U2I 링크의 SINR을 의미한다.

다음 k 번째 U2U 링크에 대한 SINR과 Throughput 수식은 식(3)부터 식(6)과 같다.

$$SINR^{u}[k] = \frac{P^{u}_{k} g_{k}}{\sigma^2 + I^{U2I}_{k} + I^{U2U}_{k}} \tag{3}$$

$$I^{U2I}_{k} = \Sigma_{m \in M} p_k[m] P^{I}_{m} \tilde{g}_{m,k} \tag{4}$$

$$I^{U2U}_{k} = \Sigma_{m \in M} \Sigma_{k' \in K\, k \neq k'} p_k[m] p_{k'}[m] P^{u}_{k'} \tilde{g}^{u}_{k',k} \tag{5}$$

$$C^{u}[k] = W \cdot log(1 + SINR^{u}[k]) \tag{6}$$

여기서 g_k는 k 번째 U2U 링크의 채널이득이다. I^{U2I}_{k}는 k 번째 U2U 링크에 대해 동일한 RB를 사용하고 있는 U2I 링크의 간섭 신호이다. I^{U2U}_{k}는 k 번째 U2U 링크에 대해 동일한 RB를 사용하는 U2U 링크의 간섭 신호이다. $\tilde{g}_{m,k}$는 m번째 U2I 링크의 송신단과 k 번째 U2U 링크의 수신단 간 채널이득이다. $p_{k'}[m]$는 k' 번째 U2U 링크의 송신단에서 m번째 RB를 사용하는지 여부이며 0 또는 1의 값을 가진다. $P^{u}_{k'}$는 k' 번째 U2U 링크의 송신단에서 전송하는 전력이다. $\tilde{g}^{u}_{k',k}$는 k' 번째 U2U 링크의 송신단과 k 번째 U2U 링크의 수신단 간 채널이득이다. $C^{u}[k]$는 k 번째 U2U 링크의 Throughput이다. 마지막으로 $SINR^{u}[k]$는 k 번째 U2U 링크의 SINR이다.

3. 심층 Q-네트워크 기반 무선 자원 할당

3.1 심층 Q-네트워크

심층 Q-네트워크는 Q-학습 알고리즘을 기반으로 하는 알고리즘이다[8]. Q-학습 알고리즘은 에이전트가 Q 값 (Q-Value, 상태–행동 쌍에 대한 누적 보상 기대 값)을 추정하는 가치 기반 강화학습이다. 누적 보상 값 수식은 식(7)과 같다.

$$R_{t:T} = r_t + \gamma r_{t+1} + \gamma^2 r_{t+1} + \ldots + \gamma^{T-1} r_{t+T+1} \tag{7}$$

여기서 r_t는 t 시점에서 환경으로부터 받은 보상 값이다. T는 환경의 마지막 시점이다. γ는 보상 감가율(Discount Factor)를 의미한다.

다음 Q 값 수식은 식(8)과 같다.

$$Q(s,a) = \mathbb{E}[R_{t:T} | s_t = s, a_t = a] \tag{8}$$

여기서 s_t는 t 시점에서 관측한 상태 값이다. a_t는 t 시점에서 선택한 행동 값이다.

Q-학습 알고리즘은 학습 알고리즘을 이용해 추정한 Q 값을 Q 테이블에 저장하는 구조이다. 하지만, Q-학습 알고리즘은 Q 테이블을 사용하기 때문에 관찰 가능한 상태 값의 수가 많아 질수록 메모리 요구량이 기하급수적으로 증가하는 한계점이 있다.

Q-학습 알고리즘과 달리 심층 Q-네트워크는 심층 신경망을 이용해 근사화한 Q 값을 구한다. 이후 $\varepsilon-greedy$ 정책에 따라 다음 행동을 선택한다. Fig. 2는 심층 Q-네트워크에서 사용하는 심층 신경망의 구조이다. 심층 Q-네트워크를 UAV 통신 시스템에 적용하기 위해서는 UAV 통신 시스템에서 활용 가능한 상태 공간, 행동 공간, 보상 함수를 필수적으로 정의해야 한다.

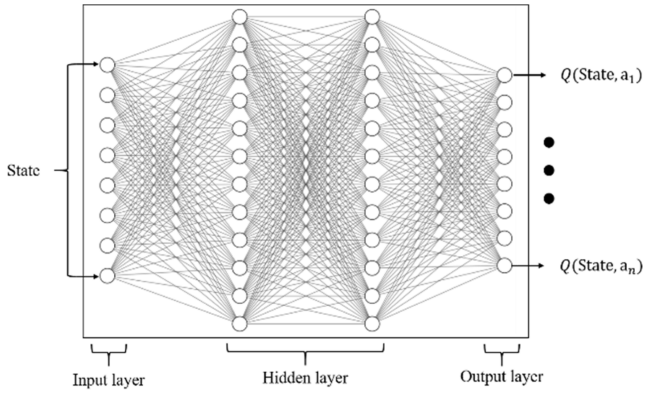

Fig. 2 The structure of Q-network

a_1	RB_1	Power $Level_1$(dBm)
a_2	RB_2	Power $Level_1$(dBm)
...		
$a_{(m\times3)-1}$	RB_{m-1}	Power $Level_3$(dBm)
$a_{(m\times3)}$	RB_m	Power $Level_3$(dBm)

Fig. 3 The example of action space

3.2 상태 공간

본 연구에서 심층 Q-네트워크를 무선 자원 할당 알고리즘에 적용하기 위해 정의한 상태 공간은 식(9)와 같다.

$$S_t = \{H_t, I_{t-1}, G_t, B_{t-1}, U_t, L_t\} \tag{9}$$

여기서 H_t는 t 시점에서 모든 U2I 링크에 대한 채널 상태 정보이다. I_{t-1}는 $t-1$ 시점에서 U2U 링크에 의해 발생한 간섭 신호 정보이다. G_t는 t 시점 U2U 링크의 송신단과 수신단 간 채널 상태 정보이다. B_{t-1}는 $t-1$ 시점에서 모든 U2U 링크에서 선택했던 RB 정보이다. U_t는 t 시점에서 U2U 링크에서 제약 시간 만기까지 남은 시간이다. L_t는 t 시점에서 요구사항 충족까지 남은 데이터 량이다.

3.3 행동 공간

본 연구에서 정의한 행동 공간은 선택 가능한 전력 레벨의 개수와 모의실험에서 형성된 RB의 개수의 조합이다.

$$a = \{a_1, a_2, a_3, \ldots a_{m\times3}\} \tag{10}$$

행동 공간은 U2U 링크의 송신단에서 선택 가능한 RB의 개수와 전력 레벨의 조합으로 구성된다. Fig. 3은 U2U 링크의 송신단의 에이전트가 선택한 행동에 따라 선택된 RB와 전송할 전력 레벨을 표현한 것이다.

3.4 보상 함수

본 연구에서 심층 Q-네트워크 기반 무선 자원 할당을 적용하기 위해 정의한 보상 함수는 식(11)부터 식(14)와 같다.

$$r_t = \lambda_{U2I} r_t^{U2I} + \lambda_{U2U} r_t^{U2U} - \lambda_p r_t^p \tag{11}$$

$$r_t^{U2I} = \frac{\Sigma_{m\in M} c_t^l[m, a_t] - \min_a \Sigma_{m\in M} C_t^l[m, a]}{\max_a \Sigma_{m\in M} C_t^l[m, a] - \min_a \Sigma_{m\in M} C_t^l[m, a]} \tag{12}$$

$$r_t^{U2U} = \frac{\Sigma_{k\in K} c_t^u[k, a_t] - \min_a \Sigma_{k\in K} C_t^u[k, a]}{\max_a \Sigma_{k\in K} C_t^u[k, a] - \min_a \Sigma_{k\in K} C_t^u[k, a]} \tag{13}$$

$$r_t^p = \frac{(T_0 - U_{t,i,j})}{T_0} \tag{14}$$

여기서 r_t^{U2I}, r_t^{U2U}, r_t^p는 순서대로 U2I 링크의 Throughput, U2U 링크의 Throughput, U2U 링크의 지연 시간 만기까지 남은 시간에 대한 보상 값이다. a_t는 t 시점에서 심층 Q-네트워크 에이전트가 선택한 행동이다. $C_t^l[m, a_t]$는 t 시점에서 선택한 행동 a_t에 따르는 m 번째 U2I 링크의 Throughput 값이다. $C_t^u[k, a_t]$는 t 시점에서 선택한 행동 a_t에 따르는 k 번째 U2U 링크의 Throughput 값이다. T_0는 미리 설정한 U2U 링크의 전송 시간 제약 값이다. $U_{t,i,j}$은 t 시점에서 i 번째 UAV에서 j 번째 UAV의 U2U 링크의 전송 시간 제약 만기까지 남은 시간이다. λ_{U2I}는 U2I 링크 보상 가중치이다. λ_{U2U}는 U2U 링크 보상 가중치이다. 마지막으로 λ_p는 U2U 링크 제약 시간 보상 가중치이다. 따라서 심층 Q-네트워크의 에이전트는 정의한 r_t^{U2I}, r_t^{U2U}, r_t^p 값에 대해서 λ_{U2I}, λ_{U2U}, λ_p 가중치 값을 고려하여 모의실험에서 계산한 누적 보상 평균 값을 최대화를 목적으로 학습하게 된다.

3.5 심층 Q-네트워크 학습 및 평가

Fig. 4는 본 연구에서 제안하는 심층 Q-네트워크의 학습 구조이다.

학습 알고리즘의 첫 번째로 Online Network와 Target Network를 초기화한다. 다음 환경으로부터 상태 값을 관측하고 Online Network를 이용해 Q 값을 계산한다. 여기서 여러 개의 UAV에서 동시에 상태를 관측하고, Q 값을 산출한다. 다음 $\varepsilon-greedy$ 정책에 따라 모든 UAV의 U2U 링크 송신단에서 다음 행동을 선택하고, 동시에 환경과 작용한다. 이후 환경으로부터 얻은 다음 상태 값과 보상 값과 함께 상태, 행동, 다음 상태, 보상 값 묶음으로 경험 재현 저장소(Replay Memory)에 저장한다. 경험 재현 저장소에 데이터 묶음이 일정 개수 이상 저장되었다면, 경험 재현 저장소에서 데이터 묶음을 무작위로 추출하여 Online Network를 학습시키는 데 사용한다. 학습시키는 데 사용한 손실 함수는 식(15)와 식(16)과 같다.

$$y = r_t + \gamma \max_a (Q(s_{t+1}, a, \theta^-)) \tag{15}$$

$$Loss(\theta) = \frac{1}{N_E} \Sigma_{s_t, a_t \in E} (y - Q(s_t, a_t, \theta))^2 \tag{16}$$

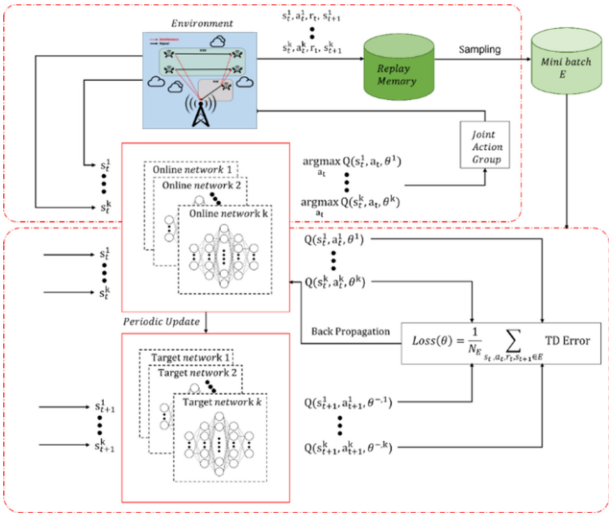

Fig. 4 The structure of DQN for the UAV communications

여기서 $Q(s_{t+1}, a, \theta^{-})$는 Target Network를 이용해 계산한 Q 값을 의미하고, $Q(s_t, a_t, \theta)$는 Online Network를 이용해 계산한 Q 값을 의미한다. E는 경험 재현 저장소에서 무작위로 추출한 데이터 묶음이다. N_E는 경험 재현 저장소에서 무작위로 추출한 데이터 묶음의 개수이다. 다음 Online Network를 일정 횟수 이상 학습하면 Online Network의 뉴런 가중치 값 θ을 Target Network의 신경망 가중치 값 θ^{-}으로 복사시킨다. 이와 같은 구조로 학습을 진행하며, 설정한 반복 횟수만큼 학습을 진행하면 학습 알고리즘을 중단시킨다.

평가 알고리즘은 학습 알고리즘과 달리 학습시킨 Online Network만을 이용한다. 첫 번째로 환경으로부터 상태 값을 관측하고 Online Network를 이용해 Q 값을 계산한다. 이후 *greedy* 정책 기반으로 다음 행동을 선택하고 환경과 상호작용을 하는 구조이다. 설정한 반복 횟수만큼 진행하면 평가 알고리즘을 중단시키고 측정한 값을 평가한다.

4. 모의실험 결과

모의실험은 Tables 1과 2와 같이 설정했다. Fig. 5 는 모의실험 환경 예시 사진이다. Fig. 6은 심층 Q-네트워크의 학습 곡선이다. 2,000번 학습 이후에 수렴하는 것을 확인할 수 있다.

모의실험은 무작위 자원 할당 기법, 채널 상태 기반 탐욕적 할당 기법과 제안하는 심층 Q-네트워크 기반 무선 자원 할당 기법을 대상으로 U2U 링크를 증가시키면서 성능을 비교했다.

무작위 할당 기법은 각 U2U 링크 송신단에서 RB를 무작위로 선택하여 23 dBm 세기 신호를 전송하는 구조이다. 채널 상태 기반 탐욕적 할당 기법은 U2U 링크 송신단에서 선택 가능한 RB에 대해서 채널 상태를 확인 후 SINR 값이 가장 높게 측정되는 RB를 선택하여 23 dBm 세기 신호를 전송하는 구조이다[9].

성능을 평가한 그래프는 3종류이며 Fig. 7은 U2U 링크의 Throughput 총합 평균, Fig. 8은 U2U 링크의 지연 시간 제약 조건 미 충족 확률, Fig. 9는 U2I 링크의 Throughput 총합 평균이다.

Table 1 Simulation parameter

Simulation size (X, Y, Height) [m]	1300 × 580 × 300
Speed of UAV [km/h]	30
Direction of UAV	Random
Path loss	Free space path loss
Shadowing [dB]	Log-normal with $\sigma^2 = 8$
Small fading	Rician fading
Noise figure σ^2 [dB]	-114
Carrier frequency [GHz]	2
Subcarrier bandwidth W [MHz]	1.5
Infrastructure antenna height [m]	25
Infrastructure antenna gain [dBi]	8
Infrastructure antenna noise figure [dB]	5
UAV antenna gain [dBi]	3
UAV antenna noise figure [dB]	9
U2U link time constraint [ms]	100
U2U link payload [Mbits]	30
Time slot [ms]	2
U2I link Transmit power level P^{I} [dBm]	23
U2U link Transmit power levels P^{u} [dBm]	5, 10, 23
Number of U2I links M	20

Table 2 Reinforcement learning parameter

Number of input layer neuron	82
Number of hidden layer neuron	520, 200, 100
Number of output layer neuron	60
Activation function of Hidden layer	Leaky ReLU
Discount factor γ	0.99
Learning rate	0.001
Coefficient of reward function [λ_{U2I}, λ_{U2U}, λ_p]	0.1, 0.3, 0.6
Number of time steps per episode	200
Number of episode	50

Fig. 7을 통해 제안하는 기법과 채널 상태 기반 탐욕적 할당 기법과 무작위 할당 기법이 거의 비슷한 성능을 볼 수 있다. 이는 제안하는 기법이 학습 과정에서 U2U 링크의 Throughput을 최대화하는 것보다 U2U 링크의 지연 시간 제약 조건을 만족시키는 것과 U2I 링크의 Throughput을 최대화하는 것을 중점으로 학습했기 때문이다. Fig. 8은 제안하는 기법이 다른 기법보다 성능이 좋은 것을 확인 할 수 있다. 여기서 채널 상태 기반 탐욕적 할당 기법은 모든 U2U 링크의 송신단에서 상태가 좋은 채널만 선택했기 때문에 간섭 신호가 크게 작용하여, 성능이 다른 기법들에 비해 떨어지는 것을 확인할 수 있다. 이는 제안하는 기법이

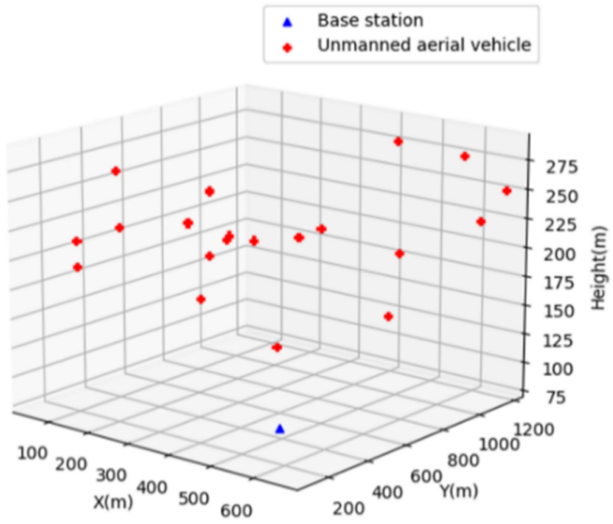

Fig. 5 Example of simulation environment

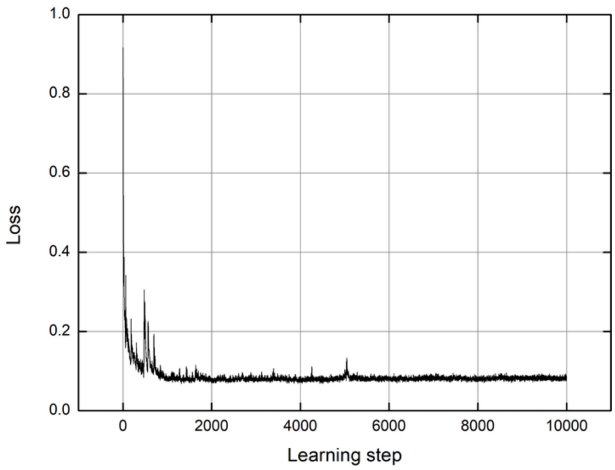

Fig. 6 Deep Q-network learning curve

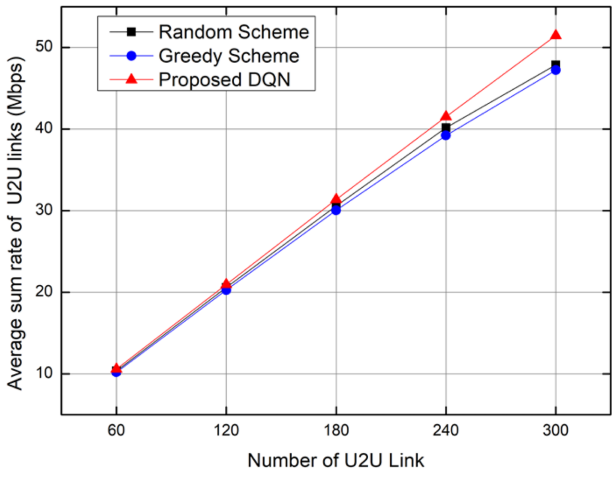

Fig. 7 Average sum rate of U2U links

다른 U2U 링크의 간섭 신호를 고려하여 행동을 선택했음을 알 수 있다.

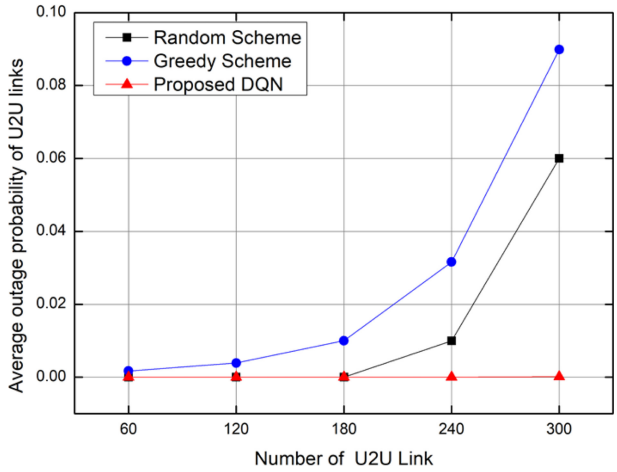

Fig. 8 Average outage probability of U2U links

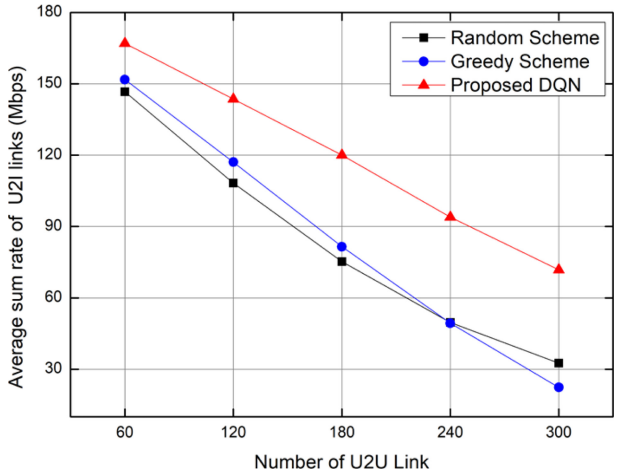

Fig. 9 Average sum rate of U2I links

마지막으로 Fig. 9 또한 제안하는 기법이 다른 기법보다 성능이 좋은 것을 확인할 수 있다. 이를 통해 제안하는 기법이 다수의 U2U 링크에서 U2I 링크의 간섭을 고려하여 자원을 행동을 선택했음을 알 수 있다.

5. 결론

본 연구를 통해 단일 기지국과 다수의 UAV가 있는 통신 환경에서 심층 Q-네트워크를 이용한 무선 자원 할당 기법의 성능의 우수함을 입증할 수 있었다. 제안하는 기법은 개별 강화 학습 에이전트가 다수의 UAV에 분산되어 개별 상태 정보를 입력받아 개별적으로 자원 할당을 하는 구조였다. 하지만 무선 자원을 효율적으로 사용하기 위해서는 다수의 UAV가 개별 상태 정보들을 공유하여 보다 협력적으로 자원을 이용할 필요가 있다. 또한 기존 심층 Q-네트워크는 이산적인 행동 공간을 사용하기 때문에 행동 공간이 증가할수록 차원의 저주(Curse of Dimensionality)에 빠지는 단점이 존재한다. 이에 따라 향후 연구로 본 연구에서 사용한

모의시험 환경에서 멀티 에이전트 강화 학습과 액터-크리틱 기반 강화 학습을 적용하는 방향으로 진행할 필요가 있다.

REFERENCES

1. Xiao, Z., Xia, P., Xia, X. G., (2016), Enabling uav cellular with millimeter-wave communication: potentials and approaches, IEEE Communications Magazine, 54(5), 66-73.
2. Bucaille, I., Héthuin, S., Munari, A., Hermenier, R., Rasheed, T., Allsopp, S., (2013), Rapidly deployable network for tactical applications: aerial base station with opportunistic links for unattended and temporary events absolute example, Proceedings of the IEEE military communications conference, 1116-1120.
3. Frew, E. W., Brown, T. X., (2008), Airborne communication networks for small unmanned aircraft systems, Proceedings of the IEEE, 2008-2027.
4. Cao, X., Yang, P., Alzenad, M., Xi, X., Wu, D., Yanikomeroglu, H., (2018), Airborne communication networks: a survey, IEEE Journal on Selected Areas in Communications, 36(9), 1907-1926.
5. Azari, M. M., Geraci, G., Garcia-Rodriguez, A., Pollin, S., (2019), Cellular uav-to-uav communications, Proceedings of the 2019 IEEE 30th Annual International Symposium on Personal, Indoor and Mobile Radio Communications, 1-7.
6. Chang Z., Guo, W., Ristaniemi, T., (2020), Machine learning-based resource allocation for multi-uav communications system, Proceedings of the 2020 IEEE International Conference on communications workshops, 1-6.
7. Yao, Z., Cheng, W., Zhang, H., (2021), Resource allocation for 5g-uav-based emergency wireless communications, International Journal of IEEE Journal on Selected Areas in Communications, 39(11), 3395-3410.
8. Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Ridemiller, M., Fidjelend, A. K., Ostrovski, G., et al., (2015), Human-level control through deep reinforcement learning, nature, 518(7540), 529-533.
9. Li, R., Zhu, P., Jin, L., (2019), Channel allocation scheme based on greedy algorithm in cognitive vehicular networks, Proceedings of the 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference, 803-807.

Dong Hee Han
M.S. Research Engineer, Unmanned/Intelligent robotics systems R&D, LIG Nex1. His research interest is in reinforcement learning for wireless communications.
E-mail: donghee.han@lignex1.com

Joo Young Kim
M.S. Chief Research Engineer, Unmanned/Intelligent robotics systems R&D, LIG Nex1. Her research interest is in machine learning-based autonomous systems.
E-mail: jooyoung.kim@lignex1.com

Sang Heung Lee
M.S. Research Engineer, Unmanned/Intelligent robotics systems R&D, LIG Nex1. His research interest is in machine learning-based autonomous systems.
E-mail: sangheung.lee@lignex1.com