불확실한 환경에서 고차원 상태 공간을 이용한 깊은 강화학습 기반의 무선 접속 방법

이태겸, 조오현*
충북대학교, 컴퓨터과학 전공
Taegyeom_l@chungbuk.ac.kr, *ohyunjo@chungbuk.ac.kr

Deep Reinforcement Learning based Medium Access using High Dimensional State Space in Dynamic and Uncertain Environments

Tae Gyeom Lee, Ohyun Jo*
Chungbuk National University, Department of Computer Science

요 약

본 논문에서는 고차원 상태 공간을 활용할 수 있는 깊은 강화학습을 기반으로 이동체를 위한 5G/6G 네트워크에서 UAV(Unmanned Aerial Vehicle)를 위한 무선 접속 기술을 제안한다. UAV는 높은 이동성을 갖기 때문에 채널 환경의 변화가 크고, 예측하기 힘든 정보만으로 통신을 시도할 확률이 높다. 기존에는 무선 접속을 위한 Back-off count 선택 시 제한적인 채널의 사용자 존재 여부에 관한 정보만을 활용하였으나 본 논문에서 제안하는 깊은 강화학습 방법론은 에이전트가 얻을 수 있는 정보만으로 최적의 행동을 선택하게 할 수 있으며, 높은 차원의 상태 데이터도 처리할 수 있다. 이러한 정보의 불확실성 문제와 대량의 정보저장 문제를 해결할 수 있다. 제안하는 강화학습 기반의 무선 접속 기술은 고차원으로 표현될 수 있는 위치 좌표와 SNR(Signal to Noise Ratio)을 이용한 Back-off 제어를 통해 네트워크 내에서 발생한 데이터 충돌 확률을 획기적으로 감소시킨다.

Ⅰ. 서 론

최근 UAV(Unmanned Aerial Vehicle)의 활용도가 높아짐에 따라 이를 적용한 응용연구가 활발히 진행되고 있다. 또한 이와 같은 이동체 네트워크에서 발생하는 문제를 효율적으로 해결하기 위해 깊은 강화학습 방법이 고려되고 있다[1]. 기존 Q-learning 모델은 상태정보를 테이블로 저장하기 때문에 고차원의 상태를 표현하려면 큰 저장 공간이 필요하다. 이러한 고차원 테이블의 또 다른 단점은 학습을 위해 매번 특정 상태정보의 위치를 찾아야하기 때문에 학습 속도가 느리다. DQN(Deep Q-learning)은 이러한 문제를 해결할 수 있다. DQN의 학습은 일정한 양의 샘플만을 저장하고 그 중 몇 개를 무작위로 뽑아 학습한 후 더 좋은 샘플을 얻으면 가장 오래된 샘플을 삭제 후 저장한다[2]. 즉, 매우 많은 수의 상태정보가 있더라도 공간의 제약 없이 학습이 가능하다. 본 논문에서는 고차원의 상태정보가 발생할 수 있는 셀룰러 네트워크에서 깊은 강화학습 모델을 활용한 UAV의 무선 접속 방법을 제안한다.

Ⅱ. 본론

1. 학습 모델

그림 1은 DQN의 학습 과정을 보여준다. UAV가 시작점에서 다시 시작점으로 돌아오기까지의 주기를 Cycle이라고 한다. UAV는 매 Frame마다 Back-off를 선택하고 전송 성공 조건에 맞게 환경으로부터 차등보상을 받는다. 또한, 자신의 위치좌표와 SNR을 포함하는 상태정보도 함께 받는다. DQN은 환경으로부터 받은 정보들을 Reply Memory의 크기만큼 저장한다. 학습이 편향적으로 되는 것을 막기 위해 Reply Memory에 저장된 정보들 중에서 무작위로 샘플들을 뽑아 학습을 진행한다. 본 논문에서 사용한 깊은 강화학습의 심층신경망은 두 개의 층으로 구성되어 있고 한 층마다 48개의 노드를 가진 Fully Connected Layer를 사용했다.

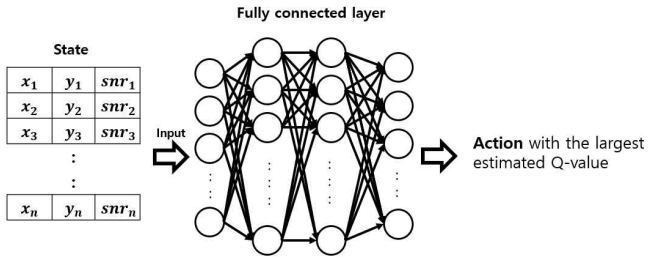

그림 1. DQN 모델 학습 과정

2. 보상 조건

강화학습에서 에이전트는 최대 보상을 목표로 행동하고 학습한다. 본 논문의 실험환경에서는 크게 3가지 조건에 따라 에이전트에 보상을 부여한다. 각 조건에 따른 보상의 최대는 +1이고 에이전트의 행동결과로 받을 수 있는 최대 보상은 +3으로 설정했다. 첫 번째 조건은 데이터 충돌의 발생이다. 다른 UAV와 같은 Back-off count를 선택한다면, 동시에 데이터를 전송하게 되므로 데이터 충돌이 발생한다. 따라서 충돌이 발생하면 -1의 보상을, 서로 다른 Back-off count를 선택해 데이터 전송에 성공하면 +1의 보상을 부여한다. 두 번째 조건은 SNR에 따른 Back-off count 선택결과이다. 본 논문의 실험에 참여하는 모든 UAV와 수신기는 자유공간에 있다. 따라서 UAV의 수신 전력(P_r)은 식 1과 같이 표현할 수 있다. λ는 반송파의 파장, P_t는 송신전력, d는 송·수신간 거리를 나타낸다.

$$P_r \propto P_t(\frac{\lambda}{4\pi d})^2 \quad (1)$$

네트워크에 UAV가 총 n개 있다면, UAV_1의 SNR(SNR_1)은 식 2와 같이 표현할 수 있다. 식 2에서 P_{r_n}은 UAV_n의 수신 전력을 의미한다.

$$SNR_1 = \frac{P_{r_1}}{P_{r_2} + P_{r_3} + ... + P_{r_n}} \quad (2)$$

따라서 SNR_1은 식 3과 같이 구할 수 있다. 식 3에서 d_n은 UAV_n과 수신기 사이의 거리를 의미한다.

$$SNR_1 = \frac{{d_2}^2 + {d_3}^2 + ... + {d_n}^2}{{d_1}^2} \quad (3)$$

SNR이 클수록 데이터 전송에 유리하므로 SNR이 큰 순서대로 빠른 Back-off count를 선택해 한정된 자원을 아낄 수 있도록 학습시키고자 했다. 조건에 맞는 순서의 Back-off count를 선택하면 +1 보상을 부여하고 그렇지 않은 경우에는 아무런 보상을 주지 않는다. 마지막은 UAV들이 선택한 Back-off count 사이의 값이 최소가 되기 위한 조건이다. 식 4와 같이 다른 UAV들이 선택한 Back-off count와 자신의 선택한 Back-off count의 차(δ)가 1이면 +1의 보상(r)을 부여한다. 차가 1보다 크면 감가변수(γ)와 -1을 곱해준 값을 부여한다. (n: Back-off count의 총 개수)

$$\begin{cases} \text{if } \delta = 1 & r = 1 \\ \text{if } \delta > 1 & r = -1 * \gamma^{n-\delta+1} \quad (\gamma = 0.9) \end{cases} \quad (4)$$

3. 실험 환경

표 1은 각 UAV의 이동반경, 통신 범위, UAV의 속도, 방향, Back-off 길이 등의 환경 파라미터를 나타낸다.

Parameters	UAV	Receiver
Radius of flight	50km	10km
Communication Range		200km
Flight speed	250km	250km
Flight direction	Clockwise/ Anticlockwise	Clockwise
Backoff length	10us	

표 1. 실험 환경 파라미터

실험의 시나리오는 [3]에서와 같이 두 대의 UAV가 정보를 수집해 한 대의 수신기에게 정보를 경쟁적으로 전달하는 시나리오이다. 각 UAV는 학습을 위해 동시에 정보 전송을 시도하며, 전송 시도 단위는 1 Frame이다. UAV는 0부터 90us 까지 10us 단위로 총 10개의 Back-off count 중 하나를 선택해 전송을 시도한다.

4. 성능 평가

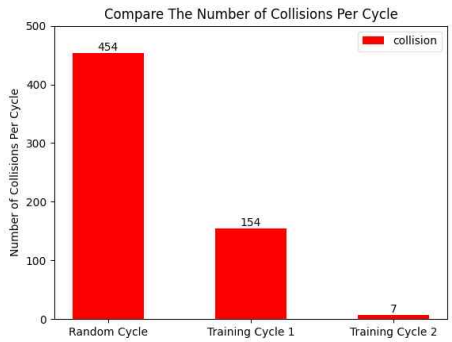

그림 2. Cycle에 따른 충돌 횟수 그래프

성능 평가는 Cycle 동안 UAV들이 받은 평균 보상을 기준으로 했으며, 네트워크 내에 데이터를 전송하는 UAV는 2대이고 각각의 항공기가 선택할 수 있는 Back-off count의 개수는 총 10개인 상황에서 평가를 진행했다. 그림 2은 학습 전과 후의 충돌횟수를 비교한 그래프이다. 학습을 하지 않은 Random Cycle동안에는 454회 정도 발생했고 2 Cycle동안 학습한 Training Cycle 2에서는 7회로 약 64배 감소한 것을 볼 수 있다. 그림 3은 보상 조건에 따라 받은 평균보상을 비교한 그래프이다. 학습 이전 SNR을 고려해 받는 보상평균은 0.44를 나타내고 있고, 학습 이후에는 약 0.35 증가했다. 학습 이전 Back-off count의 차를 고려해 받는 보상평균은 -0.19에서 학습 이후에는 최대 보상에 가까운 0.97까지 상승한 것을 볼 수 있다.

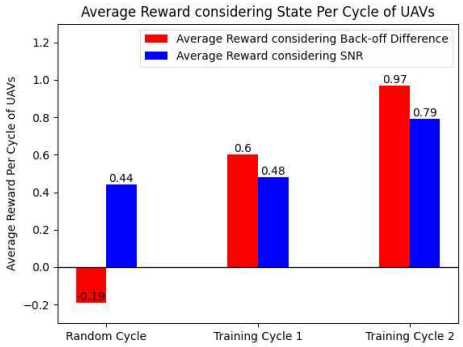

그림 3. Cycle 구간에 따른 평균 보상 그래프

Ⅲ. 결론

본 논문에서는 이동체를 위한 5G/6G 네트워크에서 UAV 간의 데이터 충돌 감소와 한정된 자원을 효율적으로 쓸 수 있도록 하는 강화학습 기반의 Back-off count 제어 방법을 제안했다. 고차원의 상태 데이터를 저장하지 않아도 되는 강화학습 방법론을 적용했고 UAV의 위치정보와 SNR 정보만을 학습해 비교적 적절한 행동을 할 수 있었다. 3가지 조건에 따라 UAV가 받은 보상은 학습 이후 최소 3배에서 6배까지 상승하는 것을 확인했고, 충돌 횟수는 대폭 감소한 것을 확인했다.

Acknowledgement

이 논문은 2020년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. 2020R1A6A1A12047945). 또한, 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 Grand ICT연구센터지원사업의 연구결과로 수행되었음(IITP-2020-0-01462). 또한, 본 과제는 행정안전부 재난안전 부처협력 사업의 지원을 받아 수행된 연구임(20008820). 또한, 이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No. 2021-0-00165, 5G+ 지능형 기지국 소프트웨어 모뎀 개발)

참 고 문 헌

[1] J. Hu, H. Zhang, L. Song, Z. Han and H. V. Poor, "Reinforcement Learning for a Cellular Internet of UAVs: Protocol Design, Trajectory Control, and Resource Management," IEEE Wireless Communications, 27, pp. 116-123, February 2020.

[2] K. Arulkumaran et al., "Deep Reinforcement Learning: A Brief Survey," IEEE Signal Processing Mag., 34, pp. 26 - 38, Nov. 2017.

[3] JungHun Byun, Sangjun Park, Joonhyeok Yoon, Yongchul Kim, Wonwoo Lee, Ohyun Jo, Taehwan Joo. Learning-Backoff based Wireless Channel Access for Tactical Airborne Networks. Journal of Convergence for Information Technology, 11, pp. 12-19, January, 2021.