

## JCR을 위한 분산 Q-learning을 통한 다중 UAV-중계기 위치 및 전력 컨트롤

박지민, 김영훈, 유희정

고려대학교 전자 및 정보공학과

{pgm8434, oh8643, heejungyu}@korea.ac.kr

Multi UAV-relay position and power control  
with Q-learning in JCR systems

Ji Min Park, Young Hoon Kim, Heejung Yu

Department of Electronics and Information Engineering, Korea University

## 요약

본 논문은 하나의 통신 파형을 이용하여 통신 기능과 레이더 기능을 동시에 사용하는 Joint Communication and Radar (JCR) 시스템이 고려되었다. JCR에서 pilot 및 data part의 전력을 조절하여 통신 성능과 레이더 성능을 최적화 할 수 있다. 또한 지상 유저에게 최적의 서비스를 제공하기 위해서는 UAV-중계기의 위치를 최적화해야 한다. 다중 UAV-중계기 환경에서는 JCR시스템의 성능 최적화를 위한 전력 조절과 UAV-중계기의 위치의 최적화가 너무 복잡해서 일반적인 최적화 문제로 해결할 수 없다. 따라서 이러한 JCR시스템 상에서 송신 파워 및 UAV-중계기의 위치를 최적화하기 위해 강화학습 접근 방식, 즉 다중 에이전트 Q-learning이 채택된다.

## I. 서론

최근 무인항공기를 이용한 다양한 무선통신 서비스가 등장하고 있다. [1] 기지국과 유저 간 가시채널 (LoS)이 확보되지 않은 경우, 무인기 중계기 (UAV-중계기)를 이용하여 유저와의 LoS를 확보하여 서비스를 제공할 수 있다. 기지국에서는 유저에게 최적의 서비스를 제공하기 위해, UAV-중계기의 위치를 파악하고 제어할 수 있어야 한다. 이러한 UAV의 위치 정보는 최적의 서비스 제공 뿐 아니라, UAV-중계기간의 핸드 오버와 같은 다양한 목적으로 사용될 수 있다.

본 연구에서는 하나의 신호를 이용하여 통신 시스템과 레이더 시스템을 사용하는 Joint Communication and Radar (JCR)을 고려하였다. JCR에서 사용되는 신호는 파일럿과 데이터 부분으로 구성된다. 파일럿 부분은 데이터 복조를 위한 채널 추정 및 UAV의 위치를 추정하는 레이더 기능에 사용된다. 데이터 부분은 지상 기지국이 UAV-중계기를 통해 유저에게 서비스를 제공할 때 데이터 전송에만 사용된다. UAV의 위치뿐 아니라 파일럿 및 데이터 구간의 전력을 조절함으로써 효율적으로 더 좋은 서비스를 제공할 수 있다. UAV-중계기의 최적화된 위치를 찾으면서 지상 기지국의 최적의 송신 전력을 조절하는 것은 간단한 최적화문제로 풀기 어렵다. 따라서 본 논문에서는 강화학습 접근 방식을 택하였으며, 분산 Q-learning을 통해 해결하였다.

## II. System model

JCR 시스템 상에서 지상 기지국이 UAV-중계기로 통신 신호를 보냄과 동시에 bistatic 레이더를 활용하여 UAV-중계기의 위치를 추적한다. 이때 사용되는 송신 신호는 파일럿과 데이터 부분으로 구성되며,  $P_{p,i}$ ,  $T_{p,i}$  및  $P_{d,i}$ ,  $T_{d,i}$ 에 의해 파일럿과 데이터 부분의 송신 전력과 길이가 결정된다.

UAV-중계기를 통한 서비스 제공에서도 환경과 UAV-중계기와 유저간의 위치에 따라 LoS채널이 확률적으로 보장된다. 이 때  $i$ 번째 UAV-중계기와 유저

간 LoS채널이 보장될 확률은 아래와 같다. [2]

$$P_{i,k}^{LoS}(\theta_{i,k}) = \frac{1}{1 + \alpha \times \exp(-\beta \times [\theta_{i,k} - \alpha])} \quad (1)$$

$$\theta_{i,k} = \frac{180}{\pi} \times \sin^{-1} \frac{h_i}{d_{i,k}} \quad (2)$$

여기서  $\theta_{i,k}$ 는 UAV-중계기와 유저 간 이루는 각도이며,  $h_i$ 는 UAV-중계기의 높이,  $d_{i,k}$ 는 UAV-중계기와 유저간의 거리,  $\alpha$ ,  $\beta$ 는 밀집된 도심 환경에서의 통계적 파라미터로 각각 12.0810, 0.1139이다.

JCR 시스템은 하나의 통신 신호를 이용하여 통신과 레이더 시스템을 동시에 사용한다. 따라서 통신 성능의 경우 Throughput을 이용하였으며, 레이더 성능의 경우 CRLB를 이용하여 평가하였다. 통신 성능의 경우 MMSE 기반의 채널 추정 방식을 이용하였으며, LoS 채널 확보 확률을 고려하여 아래와 같이 나타냈다.

$$C_i = P_{i,k}^{LoS} \frac{T_{d,i}}{T_{d,i} + T_{p,i}} B \log_2 \left( 1 + \frac{P_{d,i} \sigma_{h,i}^2}{1 + P_{d,i} \sigma_{h,i}^2} \right) \quad (3)$$

여기서  $B$ 는 송신 신호의 bandwidth이며,  $\sigma_{h,i}^2$ 과  $\sigma_{h,i}^2$ 는 각각 채널 추정치의 분산과 채널추정 에러의 분산이다.

레이더의 성능 평가의 경우 아래와 같다.

$$\sigma_{r,i}^2 \leq \frac{c^2}{32\pi^2 B_s^2 |g_i|^2 P_{p,i}} \quad (4)$$

여기서  $c$ 는 빛의 속도이며,  $B_s$ 는  $\sqrt{B}$ 이다.  $g_i$ 는 지상 기지국에서 bistatic 레이더 수신기로의 채널 이득이다.

통신 성능과 레이더 성능을 동시에 고려하기 위해 본 논문에서는 가중치의 합을 이용하여 나타냈으며, 가중치 합은 아래와 같이 정의하였다. [3]

$$U_i = w_c C_i - w_r \log_{10} \sigma_{r,i}^2 \quad (5)$$

여기서  $w_c$ 와  $w_r$ 는 각각 통신과 레이더 성능의 가중치를 의미하며, 본 논문에서는  $w_c = 1/B$ 로 고정하였으며,  $w_r$ 로 성능 우선순위를 조절하였다.

### III. 분산 Q-learning 기반 전력 및 위치 최적화 전략

Q-learning이란 마르코프 결정 과정을 통해 모델 없이 학습하는 강화학습 방법이다. Agent가 action을 선택하고, action에 따른 reward를 보상 받아 Q-table을 업데이트함을 통해 reward가 큰 방향으로 학습하는 것을 말한다.

Q-learning은 agent들이 이동할 수 있는 모든 action과 state를 고려하는 중앙 집중적 Q-learning과 각각의 agent들마다 action을 고려하여 Q-table을 업데이트 시키는 분산 Q-learning으로 나눌 수 있다. 중앙 집중적 Q-learning의 Q-table 크기는  $q^c = |A|^{N_a} \times |S|^N$ 로 나타낼 수 있으며, 여기서 A와 S는 각각 action과 state의 수,  $N_a$ 는 agent의 수를 나타낸다.

분산 Q-learning의 Q-table의 크기는  $q^d = N_a \times |A| \times |S|$ 로 나타낼 수 있다. 따라서, agent와 action의 수가 적으면 중앙 집중적 Q-learning이 분산 Q-learning에 비해 빠르게 동작하지만, agent와 action의 수가 많아지면 중앙 집중적 Q-learning의 계산 복잡성이 커지게 됨에 따라 분산 Q-learning의 효율이 더 좋아지게 된다.

본 연구에서는 분산 Q-learning 알고리즘을 제안하였으며, Q-learning에서 사용한 agent는 UAV-중계기로 가정하였다. State는 UAV-중계기의 위치와 지상 기지국의 파일럿 전력의 크기로 가정하였으며, action은 UAV-중계기의 x,y,z축의 이동과 pilot power의 상승과 감소로 나타냈었다. Reward는 UAV-중계기가 유저에게 서비스를 제공함과 동시에 JCR 시스템에서의 통신과 레이더 성능을 동시에 최대화할 수 있도록 식 (5)의 Utility function으로 정의하였으며, UAV-중계기를 통해 유저에게 서비스를 제공할 때, 성능을 최대화하는 것을 목표로 한다.

### IV. 실험 결과

앞서 제안한 Q-learning 알고리즘의 수렴과 최적성을 시뮬레이션을 통해 검증하였다. 본 시뮬레이션에서는 2대의 UAV-중계기를 가정하였으며,  $B = 2.16\text{GHz}$ ,  $w_r = 0.3$  learning rate는 0.8, discount factor는 0.7이며, initial  $\epsilon$ 은 0.99, exploration parameter는 15이며,  $T_{d,i}$ 와  $T_{p,i}$ 는 각각 999와 1이다. 지상기지국과 UAV-중계기간에는 자유공간 경로 손실 모델을 가정하였다.

제안된 Q-learning 알고리즘을 검증하기 위해 2가지 벤치마크 모델과 비교하였다. 첫 번째 벤치마크 모델은 UAV-중계기가 갈 수 있는 모든 경우를 고려하여 최대의 reward값을 계산하였다. 두 번째 벤치마크 모델은 최적의 위치에 UAV-중계기를 고정한 후, 지상 기지국의 pilot power값을 변화시켜 reward 값을 나타내었다. 제안된 Q-learning 알고리즘은 그림 1에서 볼 수 있듯 벤치마크 모델과 같이 최적의 값으로 수렴된다.

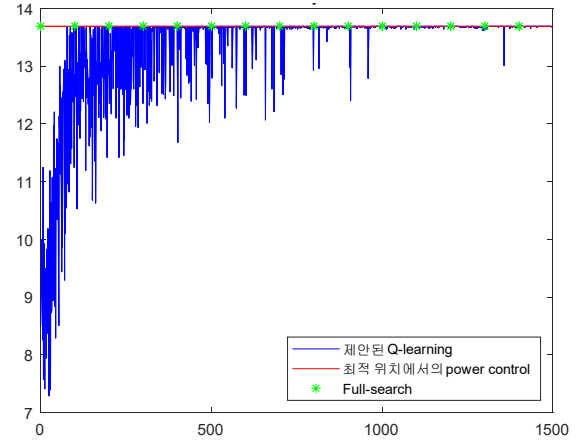


그림 1. 제안된 Q-learning 알고리즘과 벤치마크 모델의 Utility function

### V. 결론

본 논문에서는 JCR시스템 상에서 UAV-중계기를 통해 유저에게 서비스를 제공하는 환경에서 UAV-중계기의 최적의 위치와 파일럿 전력 할당을 찾아보았다. 하나의 파형을 통해 통신과 레이더 기능을 동시에 수행하는 JCR system의 성능을 평가하기 위해 throughput과 CRLB를 이용하여 Utility function을 이용하여 나타내었다. Utility function을 최적화하기 위해 UAV-중계기의 위치와 지상 기지국의 송신 파워를 조정하였으며, 최적화 문제를 풀기 위해 분산 Q-learning 알고리즘을 제안하고, 수렴 및 최적성을 검증했다.

### ACKNOWLEDGMENT

이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (2022-0-00704, 초고속 이동체 지원을 위한 3D-NET 핵심기술 개발)

### 참 고 문 헌

- [1] S. Lim, H. Yu and H. Lee, "Optimal Tethered-UAV Deployment in A2G Communication Networks: Multi-Agent Q-Learning Approach," *IEEE Internet of Things Journal*, vol. 9, no. 19, pp. 18539-18549, Oct. 2022.
- [2] S. Lee, H. Yu and H. Lee, "Multiagent Q-Learning-Based Multi-UAV Wireless Networks for Maximizing Energy Efficiency: Deployment and Power Control Strategy Design," *IEEE Internet of Things Journal*, vol. 9, no. 9, pp. 6434-6442, May, 2022.
- [3] J. M. Park, J. Cho, S. Noh and H. Yu, "Optimal Pilot and Data Power Allocation for Joint Communication-Radar Air-to-Ground Networks," *IEEE Access*, vol. 10, pp. 52336-52342, 2022.