

지상 모바일 센서로부터의 효율적 데이터 수집을 위한 심층강화학습 기반 고정익 UAV 제어 기법 연구

이승민, 전지민, 이호원
한경국립대학교 전자전기공학부
{julsin1, jimin0516, hwlee}@hknu.ac.kr

Deep Reinforcement Learning-Based Fixed-Wing UAV Control for Efficient Data Gathering from Ground Mobile Sensors

Seungmin Lee, Jimin Jeon, and Howon Lee
Hankyong National University

요약

본 논문에서는 지속적으로 움직이는 지상 센서 디바이스들로부터 효율적인 데이터 수집을 위한 고정익 UAV 제어 연구를 수행한다. 특히, 데이터 수집량과 데이터 수집 공평성을 공동으로 최대화하기 위해서 심층 강화학습 기반 고정익 UAV 제어 방안을 제안한다. 구체적으로, 실시간 학습 시 발생하는 deep Q-network(DQN)의 단점을 보완하기 위해서 dynamic n-temporal difference deep reinforcement learning (DT-DRL) 기반의 고정익 UAV 제어 기법을 제안한다. 시뮬레이션을 통해 제안하는 방안이 일반적인 DQN 방안 대비 데이터 수집량과 데이터 수집 공평성 측면에서 더 우수함을 보인다.

I. 서론

최근, 지상 디바이스의 무선 통신 수요를 효율적으로 지원하기 위해서 unmanned aerial vehicle(UAV) 기지국의 최적 제어 연구가 활발히 수행되고 있다 [1][2]. 본 연구에서는 이동성을 지닌 지상 센서 디바이스로부터의 효율적인 데이터 수집을 위한 심층 강화학습 기반 고정익 UAV 기지국 제어 연구를 수행한다. 특히, 데이터 수집량과 데이터 수집 공평성을 공동으로 최대화 하고자 한다. 이를 위해, dynamic n-temporal difference deep reinforcement learning (DT-DRL) 기반 고정익 UAV 제어 기법을 제안하며, 기존 일반적인 DQN 알고리즘 대비 데이터 수집량과 데이터 수집 공평성 성능의 우수함을 보인다.

II. 심층강화학습 기반 고정익 UAV 제어 기법

본 연구에서는 이동성을 지닌 G 대의 ground devices(GDs)가 분포된 센서 네트워크를 고려한다. 여기에서 각 GD j 는 L 사이즈의 전송해야 할 패킷을 가지고 있으며, GDs의 데이터를 수집하기 위해 한 대의 고정익 UAV가 배치되는 것으로 고려한다. 모든 GDs는 sleep 모드로 대기하며, UAV로부터 wake up signal을 받으면 active 상태가 되어 데이터 전송을 시작한다. 여기에서, 데이터 수집량과 데이터 수집 공평성을 공동으로 최대화하기 위한 UAV 제어 문제는 Markov decision process(MDP)로 설계되며, 이는 표 1에 요약되어 있다. 표1에서 \mathbf{D}^t 그리고 θ^t 는 UAV의 2차원 좌표와 이동 방위각을 나타내고, $\bar{\mathbf{D}}_j^t$ 는 GD j 의 2차원 좌표를 나타내며, $\bar{\mu}_j^t = \sum_{i=0}^t (\mu_j^i)$ 이며 $\bar{\mu}_j^t$ 와 ρ_j^t 는 t 시간까지 GD j 로부터 수집한 패킷의 양과 t 시간에서 GD j 의 잔여 패킷의 양을 나타낸다. 또한, $\Delta\theta, -\Delta\theta, 0$ 은 이동 방위각 증대/증감/유지를 나타내고, JFI 는 Jain's fairness index를 의미한다. 시간 t 에서 GD j 로부터 수집한 데이터 양은 $\mu_j^t = B_j \times \log_2(1 + SINR_j^t)$ 이다. 여기에서, B_j 는 GD j 에게 할당된 대역폭이다. DT-DRL에서 agent는 active GD를 발견할 때까지의 n-step에 대한 시간차 학습을 진행한다.

표 1. 고정익 UAV 제어를 위한 MDP 설계

State (S)	
$S^t = [\mathbf{D}^t, \theta^t, [\bar{\mathbf{D}}_1^t, \dots, \bar{\mathbf{D}}_G^t], [\bar{\mu}_1^t, \dots, \bar{\mu}_G^t], [\rho_1^t, \dots, \rho_G^t]]$	
Action (A)	Reward (R)
$A^t \in \{\Delta\theta, -\Delta\theta, 0\}$	$R^t = \sum_j^D (\mu_j^t) \times JFI$

III. 시뮬레이션 결과 및 분석

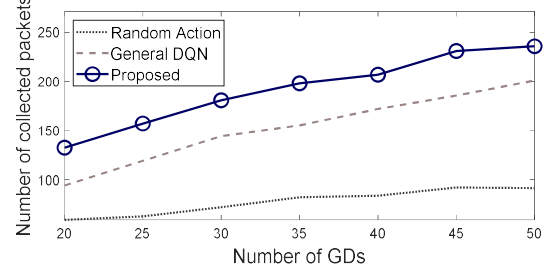


그림 1. GDs의 수에 따른 각 방안 별 데이터 수집 양

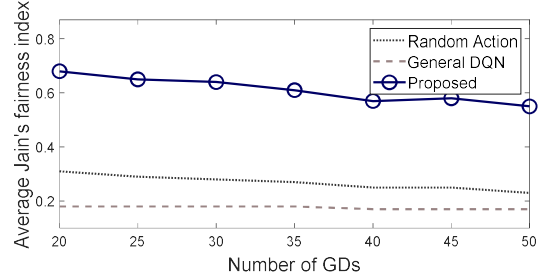


그림 2. GDs의 수에 따른 각 방안별 데이터 수집 공평성

시뮬레이션은, $L=150$ [Kbit], $B=200$ [kHz]이고 GDs의 분포가 비균일 한 환경에서 수행되었으며, 위 결과는 GDs의 수에 따른 각 방안 별 데이터 수집 량과 데이터 수집 공평성을 보여준다. 결과적으로 제안 방안은 기존 방안들 대비 데이터 수집량과 데이터 수집 공평성 측면에서 그 성능이 우수한 것을 확인할 수 있다.

ACKNOWLEDGMENT

이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(No.2022-0-00704, 초고속 이동체 지원을 위한 3D-NET 핵심 기술 개발, 50%)과 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2022R1A2C1010602, 50%).

참고 문헌

- [1] S. Lee, H. Yu and H. Lee, "Multi-agent 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.
- [2] 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.