

A2G 통합 네트워크에서 강화학습 기반 UAV 제어를 통한 지상 로봇의 분산형 임무 할당 최적화

신민규*, 조준형*, 정소이

아주대학교 AI융합네트워크학과*, 아주대학교 전자공학과

{*saycode99, *wnsgudd0126, sjung}@ajou.ac.kr

Optimization of Distributed Task Assignment for Ground Robots through Reinforcement Learning-Based UAV Control in A2G Integrated Networks

Mingyu Shin*, Junhyung Cho*, Soyi Jung

*Dept. of Artificial Intelligence Convergence Network, Ajou University,

Dept. of Electrical and Computer Engineering, Ajou University

요약

본 논문은 A2G 통합 네트워크에서 강화학습 기반의 UAV 제어를 통해 지상 로봇의 분산형 임무 할당을 최적화하는 방법을 제안한다. CBBA 알고리즘을 이용하여 지상 로봇에 임무를 할당하고, UAV를 통한 효율적인 통신 지원으로 임무 수행 효율을 극대화한다. 강화학습을 활용해 UAV의 위치를 동적으로 조정함으로써 지상 로봇이 통신 범위 내에서 원활하게 작업할 수 있도록 지원한다. 시뮬레이션을 통해 제안한 방법이 지상 로봇 간 로컬 통신만 고려했을 때 보다 더 좋은 성능을 가짐을 확인한다.

I. 서론

최근 다양한 산업 및 군사 분야에서 무인 항공기(UAV)와 지상 로봇 시스템의 통합 사용이 확대되면서 효율적인 임무 할당과 자원 최적화에 대한 연구가 필요하다. 분산 임무 할당 시스템에서는 통신 지연과 손실로 인해 로봇 간 중복 임무 할당이나 기능 손실이 발생할 수 있다. 이를 해결하기 위해 UAV를 활용한 강화학습 기반 제어 전략을 통해 지상 로봇의 임무 할당을 최적화하고자 한다. 특히, 복수 무인기 시스템의 효율적인 분산 임무 계획과 UAV 기반 하향 링크 커버리지 최적화에 초점을 맞춘다 [1][2]. 본 연구의 핵심은 CBBA(Consensus-Based Bundle Algorithm) 알고리즘을 이용해 지상 로봇에 임무를 할당하고, UAV를 통한 효율적인 통신 지원으로 임무 수행 효율을 극대화하는 것이다. 강화학습을 활용하여 UAV의 위치를 동적으로 조정함으로써 지상 로봇이 통신 범위 내에서 원활하게 작업할 수 있도록 지원한다. 이 접근 방식은 도심 및 재난 대응 환경에서 지상 로봇의 효율적인 운용을 가능하게 하며, UAV와 지상 로봇의 통합 시스템이 복잡한 환경에서도 효과적으로 작동할 수 있도록 한다. 이를 통해 모빌리티와 임무 할당의 최적화뿐만 아니라, 무인 시스템의 통합적 활용 가능성을 더욱 확장할 것으로 기대된다.

II. 시스템 모델

2.1 문제 정의

본 논문에서는 2차원 평면 내의 특정 지역에서 M 개의 임무 지점과 N 개의 지상 무인 로봇이 존재하는 상황을 다룬다. 여기서 임무 지점의 수가 지상 로봇의 수보다 많아 각 로봇의 이동 거리와 중복 임무 수행을 최소화하는 것이 목표이다. 각 지상 로봇은 일정한 속도 V_{ground} 로 이동하며, 인접한 로봇들과 지역 통신을 통해 정보를 교환한다. 지상 로봇 간의 통신 거리는 r_{ground} 로 제한된다. UAV는 지상 로봇들 간의 원활한 정보 교환

을 위해 하향 링크 통신 커버리지를 제공하며, A2G(Air-to-Ground) 채널 모델 대신 통신 거리 r_{aerial} 만을 고려한다. UAV도 일정한 속도 V_{aerial} 로 이동하며, 지상 로봇과 UAV 간의 거리가 r_{aerial} 이하일 때 지상 로봇은 UAV로부터 서비스를 받을 수 있다. 이를 통해 지상 로봇은 통신 범위 내에서 원활하게 작업이 가능하다.

2.2 지상 로봇의 분산 임무할당을 위한 CBBA 기법

CBBA는 분산된 다중 로봇 시스템에서 작업 할당 문제를 해결하는 효율적인 방법이다. 각 지상 로봇은 로컬 정보를 사용하여 작업을 할당받고, 통신을 통해 지역적인 합의를 도출하여 최종적으로 시스템의 작업 할당을 최적화한다.

CBBA는 첫 번째 번들 구성 단계에서 각 지상 로봇이 자신이 수행할 작업 목록을 번들로 구성한다.

$$c_{ij} = \begin{cases} 0, & \text{if } j \in b_i \\ \max_{n \leq |b_i|+1} S(p_i \oplus_n j) - S(p_i), & \text{otherwise} \end{cases} \quad (1)$$

각 지상 로봇 i 는 할당 가능한 작업 목록 J_i 를 가지고, 번들 b_i 와 경로 p_i 를 구성한다. 작업 j 가 번들 b_i 에 추가될 때의 한계 효용값 c_{ij} 는 수식 (1)과 같이 정의된다. 여기서 $S(p_i)$ 는 경로 p_i 의 총 보상 값을 의미하며, \oplus_n 은 작업 j 를 경로 p_i 의 n 번째 위치에 삽입하는 연산이다.

두 번째 합의 단계에서 각 지상 로봇 간 통신을 통해 번들 정보를 교환하고, 충돌을 해결한다. 지상 로봇 i 는 이웃 지상 로봇 k 와의 통신을 통해 현재의 번들 정보 (y_i, z_i, s_i) 를 교환한다. 각 작업 j 에 대해 i 는 자신과 이웃 지상 로봇의 정보를 비교하여, 업데이트 규칙에 따라 값을 갱신한다. 만약 k 가 i 보다 높은 점수를 가진다면, i 는 해당 작업 j 를 포기하고 값을 갱신한다.

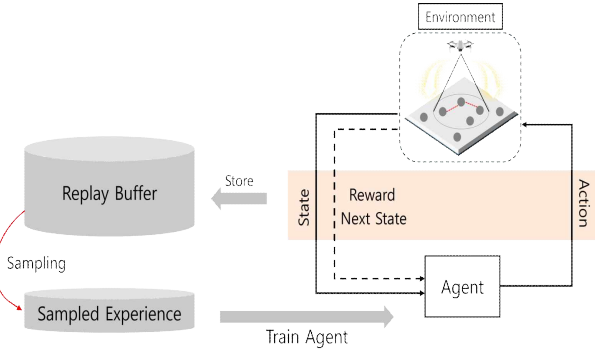


그림 1. UAV의 최적 위치 제어를 위한 심층 강화학습 Framework

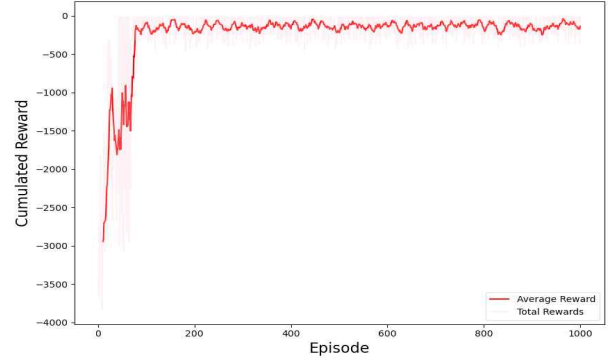


그림 2. 강화학습 결과

표 2. 강화학습 파라미터

Parameter	Value
Batch size	128
Learning rate	0.001
Discount factor	0.98
Replay buffer size	100,000

표 3. UAV 유무에 따른 CBBA 알고리즘 성능 비교 분석

성능 비교 분석 대상	UAV 배치	UAV 미배치
반복 실험 수	1000	
지상 로봇 간 통신 두절 횟수	3	11
중복 임무 수행 횟수	2	19

2.3 지상 로봇의 분산형 통신 연결성 제어

지상 로봇 간의 네트워크 연결은 수식 (2)와 같이 행렬 G 로 정의된다. 로봇 i 와 j 가 통신 연결이 되어 있을 경우 1의 값을, 그 반대의 경우 0의 값을 가진다.

$$G = \begin{bmatrix} g_{11} & g_{12} & \dots & g_{1n} \\ g_{21} & g_{22} & \dots & g_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ g_{n1} & g_{n2} & \dots & g_{nn} \end{bmatrix}, \quad (2)$$

$$g_{ij} = \begin{cases} 1 & \text{if } d_{ij} \leq r_{ground} \\ 1 & \text{if } (d_{iU} \leq r_{aerial} \text{ and } d_{jU} \leq r_{aerial}), \\ 0 & \text{otherwise} \end{cases}, \quad (3)$$

지상 로봇 간 통신 가능 여부는 수식 (3)과 같이 정의된다. 지상 로봇 i 와 j 간의 거리가 r_{ground} 이내이거나, i 와 j 가 UAV와의 거리인 r_{aerial} 이내이면 $g_{ij} = 1$ 로 정의된다. 위 두 조건 중 하나라도 만족하지 않으면 $g_{ij} = 0$ 으로 정의된다.

III. 심층 강화학습을 활용한 UAV 위치 제어 기법

본 논문에서 고려하는 문제를 MDP(Markov Decision Process)로 정의하며 UAV가 에이전트의 역할을 한다. 강화학습 기법 중 연속적인 환경에서 학습에 용이한 deep deterministic policy gradient(DDPG) 알고리즘과 강화학습 파라미터 표 2를 사용하여 강화학습의 정책을 최적화하였으며, 제안 방안의 프레임워크는 그림 1과 같다.

1) 상태(states) $S = [(x_i, z_i)_{i=1}^N, (x_u, z_u), (g_{ij})_{i,j=1}^N, (d_{ij})_{i,j=1}^N, (d_{iu})_{i=1}^N]$

로 정의한다. 각 지상 로봇 i 의 현재 위치를 $(x_i, z_i)_{i=1}^N$, UAV의 현재 위치를 x_u, z_u , 로봇 간 통신 상태를 $(g_{ij})_{i,j=1}^N$, 그리고 각 로봇 간 및 로봇과 UAV 간 거리는 $(d_{ij})_{i,j=1}^N$ 와 $(d_{iu})_{i=1}^N$ 로 나타낸다.

2) 행동(action) $A = [+ \Delta x, + \Delta z, - \Delta x, - \Delta z, 0]$ 로 정의한다. UAV는 고정된 y 좌표 상공에서 지상 로봇 간 통신 상태를 보장하는 것을 목표로 5가지 행동 중 하나의 행동을 취할 수 있다.

3) 보상(reward)은 지상 로봇 간 통신 두절 및 중복 임무 수행을 최소화하는 것을 목표로 한다. 이를 위해 모든 지상 로봇 간의 통신 가능 여부 g_{ij} 및 지상 로봇의 중복 임무 수행 횟수 C 를 기반으로 설계한다. 추가로, UAV 이동으로 인한 에너지 소모를 고려해 소량의 음의 보상을 포함한다. 임무점의 랜덤 생성에 따른 에피소드의 총 시간 변동성을 고려하여 보상을 정규화한다. 최종 보상 함수는 수식 (4)와 같이 정의된다.

$$R = \frac{1}{T} * \left(\sum_{t=1}^T \sum_{j=1}^N \sum_{i=i+1}^N -1 * (1 - g_{ij}(t)) \right) + (-0.5 * \frac{C}{T}) + (-0.1 * \frac{S}{T}), \quad (4)$$

IV. 시뮬레이션 결과 분석 및 고찰

시뮬레이션은 $N = 12$, $M = 4$, $V_{ground} = 20[m/s]$, $V_{aerial} = 80[m/s]$, $r_{ground} = 500[m]$, $r_{aerial} = 2000[m]$ 인 환경에서 수행되었으며, 임무 지점은 $7000[m] \times 7000[m]$ 크기의 지상 환경에 랜덤하게 분포된다. 한 에피소드 내에서 $t = 100[s]$ 일 때 임무 재할당이 이루어지고, 재할당된 임무 지점을 모두 도달했을 때 에피소드는 종료된다. 그림 2에서 학습 진행에 따른 누적 보상이 수렴하는 것을 확인할 수 있다. 또한, 표 3에서 제안한 방법이 지상 로봇 간 로컬 통신만 고려한 임무 할당보다 성능이 더 효과적임을 확인할 수 있다.

ACKNOWLEDGMENT

이 연구는 2023년도 산업통상자원부 및 산업기술평가관리원(KEIT) 연구비 지원에 의한 연구임(과제번호: 1415187715)

참고 문헌

[1] H.-L. Choi, L. Brunet, and J. P. How, "Consensus-based decentralized auctions for robust task allocation," *IEEE Transactions on Robotics*, vol. 25, no. 4, pp. 912-926, Aug. 2009.

[2] S. Lim, H. Yu and H. Lee, "Optimal Tethered-UAV Deployment in A2G Communication Networks: Multi-Agent Q-Learning Approach," in *IEEE Internet of Things Journal*, vol. 9, no. 19, pp. 18539-18549, Oct. 2022.