

부분 관찰 다중 에이전트 강화 학습 기반 무인 항공기 모니터링 시스템을 위한 시맨틱 통신 기술 연구

상천가, 최민석*
경희대학교, *경희대학교

singskyx@khu.ac.kr, *choims@khu.ac.kr

Semantic Communications for Partially Observable Multi-Agent Reinforcement Learning-Based Unmanned Aerial Vehicles Monitoring System

Tiange Xiang and Minseok Choi
Kyung Hee University

요약

다중 에이전트 강화 학습(MARL)에서 에이전트 간 통신은 글로벌 최적 행동을 위해 중요하다. 하지만 불안정한 통신 환경에서는 정보 공유 효과가 감소한다. 이를 해결하기 위해 본 논문은 시맨틱 통신 기술을 제안하며, LSTM 및 self-attention을 사용한 시맨틱 인코더와 디코더를 설계하였다. 이 접근법은 UAV 모니터링 시스템에 적용되어 재난 지역을 효과적으로 모니터링하는 데 유용함을 시뮬레이션으로 검증하였다. 결과적으로, 불안정한 통신 환경에서 시맨틱 통신이 Age of Information (AoI) 성능을 향상시킴을 확인하였다.

I. 서론

현실 세계에서 다중 에이전트 강화 학습(MARL)은 자율 주행[1], 로봇[2]과 같은 응용 분야에서 복잡한 상호 작용을 해결하는 데 매우 중요하다. 이러한 시스템은 부분 관찰성, 환경 변화, 에이전트 전략의 non-stationarity[3]로 인해 효과적인 에이전트 간 통신이 필요하다. 재난 모니터링 및 응급 관리에서 무인 항공기(UAV)의 배치는 기동성과 적응력 때문에 증가하고 있다. 그러나 재난 환경의 복잡성과 전송 지연, 패킷 손실과 같은 불안정한 통신 조건은 효과적인 UAV 협업에 큰 어려움이다 [4]. 이때, 시맨틱 통신은 원시 관찰 데이터를 압축하고 과거의 데이터까지 활용하여 정보 교환의 효율성과 정확성을 높여 다중 UAV 시스템을 강화할 수 있다. 본 논문은 재난 시나리오에서 MARL 기반 UAV 모니터링 시스템을 위한 시맨틱 통신 방식을 제안한다.

II. 다중 무인기 모니터링 시스템

본 논문에서는 $D \times D$ 영역 내 가우시안 분포로 무작위 생성된 N 대의 UAV와 K 개의 유한한 범위를 갖는 재난 현장을 고려한다. 재난 현장 i 의 주변 좌표 (x, y) 의 중요도는 다음과 같이 가우시안 확률 분포로 모델링된다.

$$P_i(x, y, t) = \frac{1}{2\pi\sigma_{i,x}\sigma_{i,y}} \exp\left(-\frac{(x-\mu_{i,x}(t))^2}{2\sigma_{i,x}^2} - \frac{(y-\mu_{i,y}(t))^2}{2\sigma_{i,y}^2}\right) \quad (1)$$

여기서 $\mu_{i,x}(t)$ 와 $\mu_{i,y}(t)$ 는 재난 지역 i 의 중심을 나타내며, $\sigma_{i,x}$ 와 $\sigma_{i,y}$ 는 표준 편차를 나타낸다. 이 모델은 재난 중심에 가까워질수록 관측 중요도가 증가하는 실제 시나리오를 반영하며, 다양한 크기와 형태의 영역을 수용한다. 재난 지역은 속도 v 로 천천히 이동하며, 일정 시간 간격 Δt 마다 다음과 같이 중심을 업데이트한다.

$$\mu_i(t + \Delta t) = \mu_i(t) + v \cdot \Delta t \quad (2)$$

각 UAV는 제한된 반경 R 내 좌표를 관측하며, 제한된

데이터를 기반으로 이동을 결정해야 한다. 이때 이미 발견된 지역을 계속 모니터링하거나 새로운 지역을 탐색할 수 있다. 좁은 관측 범위 때문에 UAV는 전체 상태 정보를 수집할 수 없고, 더 나은 결정을 위해 제한된 범위의 통신 링크로 다른 UAV와 데이터를 공유해야 한다. 각 UAV는 본인이 관측한 정보와 다른 UAV로부터 수신한 정보를 바탕으로 {위, 아래, 왼쪽, 오른쪽, 대기} 중에서 최적의 행동을 선택한다. 재난 지역의 역동적인 특성을 고려하여, 관측의 freshness를 측정하는 Age of Information (AoI) 개념을 도입한다. 시간 t 에 재난 지역 i 의 좌표 (x, y) 에 대한 AoI는 다음과 같이 정의된다.

$$\delta_i(x, y, t) = \begin{cases} \delta_i(x, y, t-1) + 1 & \text{if } (x, y) \notin S_n(t), \forall n \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

여기서 $S_n(t)$ 는 시간 t 에 UAV n 이 관측할 수 있는 좌표의 집합이다. 모니터링 성능은 평균 AoI (AAoI)와 가중 평균 AoI (WAoI)를 계산하여 평가된다:

$$\delta_i(t) = \frac{1}{|G_i|} \sum_{(x,y) \in G_i} \delta_i(x, y, t) \quad (4)$$

$$w_i(t) = \frac{1}{|G_i|} \sum_{(x,y) \in G_i} P_i(x, y, t) \delta_i(x, y, t) \quad (5)$$

우리의 다중 UAV 모니터링 시스템은 $\delta_i(t)$ 와 $w_i(t)$ 값을 낮게 유지하여 관측된 정보가 가능한 한 freshness 하도록 유지하는 것을 목표로 한다.

III. 다중 UAV 모니터링 시스템을 위한 MARL 프레임워크

각 UAV n 은 시간 t 에 주변 정보를 부분적으로 관찰하며, $a_n(t)$ 로 표시되며, 이는 기하학적 관찰 $g_n(t)$ (지도 이미지 $g_n^m(t)$ 및 AoI $g_n^a(t)$)와 벡터 관찰 $v_n(t)$ (재난 장면의 중심 위치 및 다른 UAV의 상대 위치)로 구성된다.

글로벌 상태 $s(t)$ 는 모든 에이전트의 관찰을 합친 것이며, 지도 이미지와 AoI 값을 포함한다. 여러 UAV가 동일한 위치를 관찰할 경우, 전역 상태는 더 낮은 AoI 값을 채택한다. $o_n(t)$ 를 기반으로 UAV n 은 $A=\{\text{위, 아래, 오른쪽, 왼쪽, 대기}\}$ 에서 다음 이동 방향을 결정하는 행동 $a_n(t)$ 를 취한다. 보상 함수는 중요한 지역을 우선적으로 모니터링하고 관찰 데이터의 신선도를 유지하기 위해 정의된다:

$$R(s, a, t) = -\sum_{i=1}^K \sum_{(x,y) \in S(t)} P_i(x, y, t) \delta_i(x, y, t) \quad (6)$$

여기서 $S(t) \triangleq \cup_{n=1}^N S_n(t)$ 이 보상은 각 에이전트별로 분리 가능하다:

$$R(s, a, t) = \sum_{n=1}^N R_n(o_n, a_n, t) = -\sum_{n=1}^N \sum_{i=1}^K \sum_{(x,y) \in S_n(t)} \frac{1}{\tilde{n}(x,y,t)} P_i(x, y, t) \delta_i(x, y, t) \quad (7)$$

여기서 $\tilde{n}(x, y, t)$ 는 동일한 시간 t 에 (x, y) 를 관찰하는 UAV의 수이다. 각 UAV는 $v_n(t)$ 를 사용하여 중복 관찰을 계산할 수 있다. 에이전트 n 는 보상 $R_n = \sum_{t=0}^T \gamma^t R_n(o_n, a_n, t)$ 를 최대화하는 것을 목표로 하며, 여기서 $\gamma \in [0, 1]$ 이다. 우리는 다중 에이전트 Actor-Critic 구조를 기반으로 정책을 훈련하여 할인된 보상을 최대화한다.

IV. 실험 결과

100,000 회 반복 학습의 보상함수곡선은 송신 전력 $P_t=70W$ 와 $P_t=25W$ 로 각각 Fig.1-2에 나타난다. $P_t=70W$ 에서는 높은 SNR을 유지해 관찰 데이터를 거의 전송 실패 없다. 보상 함수가 빠르게 수렴한다. 반면, $P_t=25W$ 에서는 낮은 SNR로 전송 실패가 증가하지만, 시맨틱 통신은 빠르게 수렴한다. LSTM 및 self-attention 인코딩/디코딩이 중요한 역할을 하며, 시맨틱 통신은 전송 실패를 줄이고 보상 함수의 수렴 속도를 높인다. 따라서 시맨틱 메시지 인코딩 방법이 필수적이다.

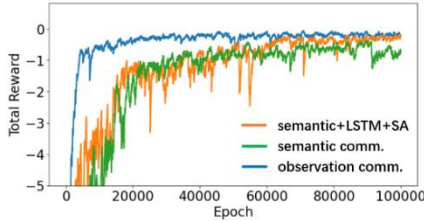


Fig.1 Training Progress with $P_t = 70 W$

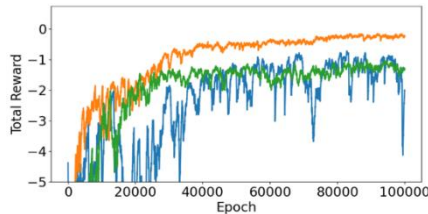


Fig.2 Training Progress with $P_t = 25 W$

본 논문에서는 다중 UAV 모니터링 시스템에서 시맨틱 통신 기반 MARL 프레임워크와 시맨틱 인코딩/디코딩 방법의 성능을 검증하였다. 다양한 송신 전력 수준에서 비교하였으며, 목표는 재난 지역을 지속적으로 모니터링하여 데이터를 최신 상태로 유지하는 것이다. 평균 WAoI를 성능 지표로 사용하였다. 높은 SNR($P_t=70W$)에서는 기술들 간 큰 성능 차이가 없었으며, 낮은 SNR($P_t=25W$)에서는 성능 차이가 커졌다. LSTM 및 self-attention 기반 시맨틱 인코딩/디코딩은 낮은 WAoI를 유지하여 통신 효율성과 정보 신선도를 향상시켰다.

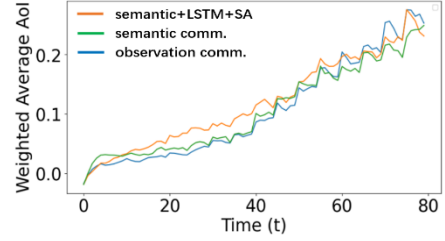


Fig.3 Disaster Scenes, $P_t = 70 W$

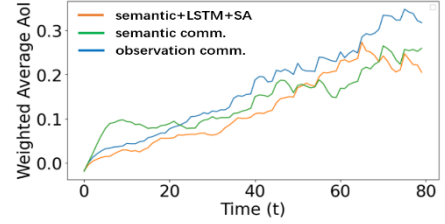


Fig.4 Disaster Scenes, $P_t = 25 W$

V. 결론

본 연구에서는 불완전한 통신 환경에서 MARL 시스템의 성능 향상을 위해 시맨틱 통신 기술을 도입하였다. 다중 UAV 모니터링을 위해 시맨틱 통신 지원 Actor-Critic 프레임워크를 설계하고, 수정된 ResNet-18, LSTM, self-attention 네트워크를 사용한 인코더 및 디코더 아키텍처를 개발하였다. 시뮬레이션 결과, 제안된 방식은 낮은 SNR에서도 성능(보상 함수, 평균 WAoI)이 크게 저하되지 않았다.

ACKNOWLEDGMENT

이 논문은 2024년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원과 한국연구재단의 지원을 받아 수행된 연구임 (No. 2021-0-02201, 사용자 프라이버시를 보존하는 비디오 캐싱을 위한 연합 학습 시스템, NRF-2022R1C1C1010766, No. 2022R1A4A3033401)

참고 문헌

- [1] Shalev-Shwartz, Shai et al. "Safe, Multi-Agent, Reinforcement Learning for Autonomous Driving." *ArXiv abs/1610.03295* (2016): n. pag
- [2] Jens Kober, J Andrew Bagnell, and Jan Peters. Reinforcement learning in robotics: A survey. *The International Journal of Robotics Research*, 32(11):1238-1274, 2013
- [3] Georgios Papoudakis, Filippos Christianos, Arrasy Rahman, and Stefano V Albrecht. Dealing with non-stationarity in multi-agent deep reinforcement learning. *arXiv preprint arXiv:1906.04737*, 2019
- [4] Communication enabled deep reinforcement learning to optimize energy efficiency in uav ssisted networks. *Vehicular Communications*, 43:100640, 2023