

심층 강화학습 기반 UAV 통신 시스템의 상향링크 성능 및 에너지 효율 최적화

엄준수, 신오순
숭실대학교

jseom@soongsil.ac.kr, osshin@ssu.ac.kr

Uplink Performance and Energy Efficiency Optimization in UAV Communication Systems Based on Deep Reinforcement Learning

Joon-Soo Eom, Oh-Soon Shin
Soongsil University

요약

본 연구에서는 다수의 시간 단계에 걸친 무인항공기(UAV) 지원 통신 네트워크에서 UAV의 이동 에너지 소비를 최소화하는 동시에 상향링크 최소 전송률을 최대화하기 위한 새로운 심층 강화학습 기반 최적화 기법을 제안한다. 모의실험 결과를 통해 제안된 기법이 상향링크 통신 성능과 에너지 효율을 공동으로 고려한 최적화 문제에서 효과적이고 장인한 성능을 달성함을 확인하였다.

I. 서론

무인 항공기(UAV)는 우수한 이동성과 자율 운용 능력을 바탕으로 긴급 상황이나 통신 소외 지역에 신속하게 배치되어 유연한 커버리지를 제공할 수 있어, 6G 통합 네트워크의 핵심 구현 요소로 주목받고 있다. 그러나 UAV에 탑재된 에너지 용량은 제한적이어서, 에너지 효율을 고려한 운용은 UAV 통신 시스템 설계에서 여전히 중요한 연구 과제로 남아 있다[1]. 본 연구에서는 DDPG(Deep Deterministic Policy Gradient) 알고리즘을 적용하여 상향링크 성능과 에너지 효율을 공동으로 최적화하고, 이를 통해 국소 최적해를 도출하고자 한다.

II. 시스템 모델 및 제안하는 알고리즘

UAV의 높은 고도로 인해 지상 사용자로부터의 상향링크 채널 \mathbf{h}_u^{ul} 은 LoS 성분이 지배적인 환경으로 가정하며, 대규모 경로 순서와 소규모 페이딩을 모두 고려하여 모델링한다. 시간 슬롯 $s \in \mathcal{S} = \{1, 2, \dots, S\}$ 의 s 번째 단계에서 상향링크 사용자 UE_u^{ul}의 신호 대 간섭 및 잡음비(Signal-to-Interference-plus-Noise Ratio: SINR)는 $\gamma_u^{\text{ul}}(s)$ 로 정의된다. 이에 대응하는 상향링크 전송률은 다음과 같이 표현된다.

$$R_u^{\text{ul}}(s) = \eta_u(s) B_{\text{up}} \log_2 (1 + \gamma_u^{\text{ul}}(s)) = \eta_u B_{\text{up}} \log_2 \left(1 + \frac{|\mathbf{h}_u^{\text{ul}}(s)|^2}{\sigma^2} \right) \quad (1)$$

여기서 $B_{\text{up}} = 10 \text{ MHz}$ 는 전체 상향링크 대역폭을 의미하며, $\eta_u(s) \in \{0, 1\}$ 는 주파수 분할 다중 접속(FDMA) 환경에서 사용자 u 에게 대역폭이 할당되었는지를 나타내는 이진 변수이다.

UAV의 에너지 소비를 모델링하기 위해, 본 연구에서는 [2]에서 제시된 Fixed-wing UAV에 대한 해석적 에너지 모델을 채택하였다. s 번째 시간 단계에서 UAV의 위치와 속도는 각각 $\mathbf{u}(s) = (x(s), y(s), z(s))$ 와 $\mathbf{v}(s)$ 로 표현되며, 이에 대한 운동학적 관계는 $\mathbf{u}(s+1) = \mathbf{u}(s) + \mathbf{v}(s)\delta_t + \frac{1}{2}\boldsymbol{\omega}(s)\delta_t^2$ 와 $\mathbf{v}(s+1) = \mathbf{v}(s) + \boldsymbol{\omega}(s)\delta_t$ 로 주어진다. 여기서 $\boldsymbol{\omega}(s)$ 는 시간 슬롯 s 에서의 UAV의 가속도를 의미한다. Fixed-wing UAV의 추진 전력을

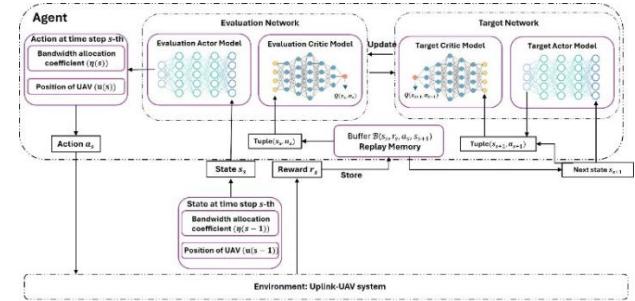


그림 1. 상향링크 UAV 시스템을 위한 제안된 DDPG 기반 알고리즘 구조의 블록도.

시간 슬롯 s 에서 $\hat{p}(s) = a \|\mathbf{v}(s)\|^3 + \frac{b}{\|\mathbf{v}(s)\|} \left(1 + \frac{\|\boldsymbol{\omega}(s)\|^2}{g^2} \right)$ 과 같이 정의된다. 여기서 g 는 중력 가속도를 나타내며, a 와 b 는 각각 UAV의 날개 면적, 기체 중량 및 공기 밀도와 관련된 상수 파라미터이다. 이를 바탕으로 UAV의 이동에 따른 시간 슬롯 s 에서의 에너지 소비는 $E_{\text{uav}}^{\text{mov}}(s) = (p(s) + \hat{p}(s))$ 와 같이 모델링된다.

본 연구의 목적은 S 개의 시간 단계에 동안 K_{ul} 명의 상향링크 사용자에 대해 최소 상향링크 전송률을 최대화하는 동시에, UAV의 이동 에너지를 최소화하는 것이다. 이에 따라 최적화 문제를 다음과 같이 수립한다.

$$\max_{\eta, \mathbf{u}} \sum_{s=1}^S \max_u (R_u^{\text{ul}}(s) - \epsilon_0 \sum_{s=1}^S E_{\text{uav}}^{\text{mov}}) \quad (2a)$$

$$\text{s.t. } 0 \leq \sum_{u=1}^{K_{\text{ul}}} \eta_u(s) \leq 1, \forall s \in \mathcal{S}, \quad (2b)$$

$$\|\mathbf{u}(s) - \mathbf{u}(s-1)\| \leq L_{\text{max}}, \forall s \in \mathcal{S}, \quad (2c)$$

$$h_{\min} \leq z_{\text{uav}}(s) \leq h_{\max}, \forall s \in \mathcal{S}. \quad (2d)$$

여기서 스칼라 값 ϵ_0 은 상향링크 성능과 에너지 효율 간의 절충을 조절하는 가중치이다. 제약식 (1b)은 대역폭 예산을 보장하며, 제약식 (1c)은 시간 단계당 UAV의 이동 거리를 제한하고, 제약식 (1d)은 UAV의 고도를 허용 가능한 운용 범위 내로 제한한다.

문제 (1)은 비선형성 및 연속적인 제어 변수를 포함하는 비블록 최적화 문제이므로, 이를 해결하기 위해 연속적인 행동 공간에 적합한 Model-free Off-policy Actor-critic 심층 강화학습 기법인 DDPG 알고리즘을 적용하였다. DDPG는 그림 1과 같이 네 개의 신경망, 즉

알고리즘 1. 제안된 DDPG 기반 알고리즘

- 1: 초기화(Initialization):
 $Q(\mathbf{s}_s, \mathbf{a}_s, \theta^q), \mu(\mathbf{s}_s, \mathbf{a}_s, \theta^\mu), Q'(\mathbf{s}_s, \mathbf{a}_s, \theta^{q'}), \mu'(\mathbf{s}_s, \mathbf{a}_s, \theta^{\mu'})$ 를 초기화
이때 $\theta^{q'} = \theta^q, \theta^{\mu'} = \theta^\mu$ 로 설정
 $B, \epsilon_1, \epsilon_2, \xi, N_B, \text{ep}_{\max}$ 를 설정
- 2: Episode 인덱스 $\text{ep} \leftarrow 1$, 시간 단계 $s \leftarrow 1$ 로 설정
- 3: $\text{ep} = 1$ 부터 ep_{\max} 까지 반복
 - 4: 시간 단계 0에서 UAV 위치 $\mathbf{u}(0)$ 을 설정
 - 5: 초기 상태 \mathbf{s}_1 을 설정
 - 6: $s = 1$ 부터 S 까지 반복
 - 7: $\mathbf{a}_s = \mu(\mathbf{s}_s | \theta^\mu) + \mathcal{N}_s$ 선택
 - 8: $\mathbf{u}(s)$ 와 $\eta(s)$ 를 추출
 - 9: $R_u^{\text{ul}}(s)$ 와 $E_{\text{uav}}^{\text{mov}}(s)$ 를 계산
 - 10: 다음 상태 \mathbf{s}_{s+1} 을 설정
 - 11: 보상 r_s 를 계산
 - 12: $(\mathbf{s}_s, \mathbf{a}_s, r_s, \mathbf{s}_{s+1})$ 를 \mathcal{B} 에 저장
 - 13: \mathcal{B} 에서 N_B 만큼 무작위 샘플링
 - 14: Update Evaluation Critic Network
 - 15: Update Evaluation Actor Network
 - 16: Update Target Networks
 - 17: 종료
 - 18: 종료

Actor μ , Target Actor μ' , Critic Q , Target Critic Q' 과 이들에 대응하는 파라미터 $\theta_\mu, \theta_q, \theta'_\mu, \theta'_q$ 로 구성되며, 결정론적 정책 경사 이론에 기반하여 경사 하강법을 통해 학습된다[3]. DDPG 기반 최적화 절차는 알고리즘 1에 상세히 제시되어 있으며, 주요 Hyperparameter는 다음과 같다: Soft Update 계수 $\tau = 0.005$, Learning Rates $\epsilon_1 = 0.001, \epsilon_2 = 0.002$, Discount Factor $\xi = 0.99$, Replay Buffer Size $B = 10^6$, Mini-batch Size $N_B = 256$. 각 시간 단계 s 에서 상태는 $\mathbf{s}_s = \{\mathbf{u}(s-1), \mathbf{a}^{\text{ul}}, (s-1)\}$ 로 정의되며, 행동은 $\mathbf{a}_s = \{\mathbf{u}(s), \eta(s)\}$ 로 구성된다. 보상 함수는 $r_s = \min_u (R_u^{\text{ul}}(s)) - \epsilon_0 E_{\text{uav}}^{\text{mov}}(s)$ 으로 정의되며, 이는 상향링크 전송률과 UAV 에너지 소비 간의 균형을 반영한다. Actor 및 Critic 신경망은 Replay Buffer에서 Sampling 된 Mini-batch를 이용하여 반복적으로 업데이트됨으로써 학습 안정성과 수렴 성능을 향상시킨다.

III. 모의실험 결과

제안된 강화학습 기반 접근법은 학습 초기 단계부터 우수한 수렴 특성을 보이며, Episode 수가 증가함에 따라 누적 보상이 점진적으로 증가한 후 안정적으로 수렴하는 경향을 나타낸다. 특히 약 1,000 Episode 이내에 준최적 수준의 누적 보상에 도달함으로써, 제안된 알고리즘이 비교적 짧은 학습 시간 내에 효과적인 제어 정책을 학습할 수 있음을 확인할 수 있다. 이러한 결과는 본 논문에서 수립한 비볼록 최적화 문제를 해결하는 데 있어 제안된 학습 기법이 높은 학습 효율성과 안전성을 동시에 갖추고 있음을 시사한다.

또한 가중 계수 ϵ_0 는 상향링크 성능과 UAV 이동 에너지 소비 간의 상충 관계를 조절하는 핵심 파라미터로 작용한다. 그림 2에서 확인할 수 있듯이, ϵ_0 값이 증가할수록 에너지 절감이 우선적으로 반영되어 UAV의 이동 에너지 소비는 감소하는 반면, 상향링크

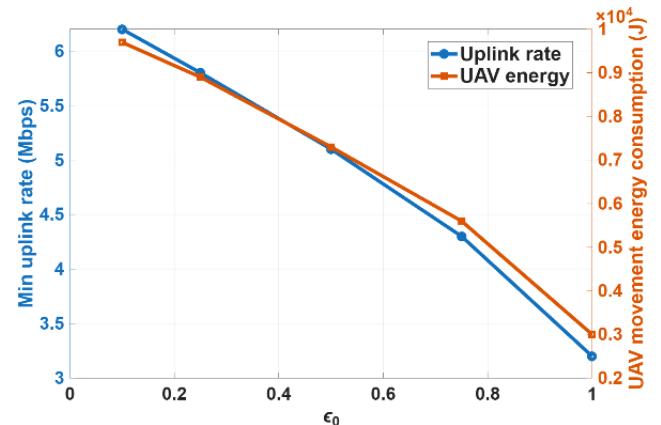


그림 2. 가중 계수 ϵ_0 에 따른 상향링크 성능과 UAV 에너지 소비 간의 Tradeoff.

전송률을 포함한 통신 성능은 점진적으로 저하되는 경향을 보인다. 이는 에너지 효율을 과도하게 강조할 경우 통신 품질이 희생될 수 있음을 의미한다. 따라서 시스템의 서비스 요구 사항과 운용 환경을 고려하여 통신 성능과 에너지 효율 간의 적절한 절충을 달성할 수 있도록 ϵ_0 를 신중하게 설정하는 것이 중요하다.

IV. 결론

본 연구에서는 상향링크 통신 성능과 UAV의 이동 에너지 소비를 공동으로 고려한 DRL 기반 최적화 알고리즘을 제안하였다. 모의실험 결과를 통해 제안된 알고리즘이 짧은 학습 시간 내에 안정적으로 수렴하며, 통신 성능과 에너지 효율 간의 효과적인 절충을 달성할 수 있음을 확인하였다. 향후 연구에서는 사용자 간 간섭을 명시적으로 고려한 모델링, 다수의 UAV 가 협력적으로 운용되는 시나리오, 그리고 보다 현실적인 채널 및 에너지 모델을 포함하는 방향으로 본 연구를 확장할 계획이다.

ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. RS-2025-02214082).

참고문헌

- [1] J. Mu, R. Zhang, Y. Cui, N. Gao and X. Jing, "UAV meets integrated sensing and communication: Challenges and future directions," *IEEE Commun. Mag.*, vol. 61, no. 5, pp. 62–67, May 2023.
- [2] Y. Liu, S. Liu, X. Liu, Z. Liu and T. S. Durrani, "Sensing fairness-based energy efficiency optimization for UAV enabled integrated sensing and communication," *IEEE Wirel. Commun. Lett.*, vol. 12, no. 10, pp. 1702–1706, Oct. 2023.
- [3] D. Silver *et al.*, "Deterministic policy gradient algorithms," in *Proc. 31st Inter. Conf. Mach. Learn. (ICML)*, vol. 32, pp. 387–395.