

강건한 안티드론 방공 시스템을 위한 LLM-유도 강화학습 기반 무인기 자율 제어

김규선, 김중헌

고려대학교

kingdom0545@korea.ac.kr, joongheon@korea.ac.kr

LLM-based Reinforcement Learning for UAV Control in Defense Systems

Gyu Seon Kim and Joongheon Kim

Korea University

요약

최근의 무력 분쟁은 현대 전장에서 무인항공기(Unmanned Aerial Vehicles, UAV)가 수행하는 결정적 역할을 분명히 부각시켰으며, 특히 저비용 군집 드론이 기존의 미사일 기반 방공 체계를 무력화할 수 있는 능력을 입증하였다. 저가의 적대 드론을 고비용 요격 수단으로 대응해야 하는 상황에서 발생하는 경제적·전술적 비대칭성은, 적응적이며 비용 효율적인 안티드론 방어 시스템(Anti-Drone Defense Systems, ADS)으로의 패러다임 전환을 요구한다. 비례-적분-미분 제어 및 모델 예측 제어를 포함한 기존 제어 기법들은 예측 불가능하고 동적으로 변화하는 환경에서의 적응성이 제한적이라는 한계를 지닌다. 강화학습(Reinforcement Learning, RL)은 자기 최적화 정책 학습을 통해 대안이 될 수 있으나, 느린 수렴 속도로 인해 실시간 전장 환경에서 요구되는 신속한 의사결정을 심각하게 저해한다. 이러한 문제를 해결하기 위해 본 연구에서는 대규모 언어 모델(Large Language Model, LLM)의 추론, 추상화, 계획 수립 능력을 강화학습 학습 루프에 통합한 LLM-유도 강화학습 기반 안티드론 시스템(Large Language Model - Guided Reinforcement Learning - based Anti-Drone System, LRL-ADS)을 제안한다. 제안된 프레임워크는 LLM이 생성한 사전 정보를 정책 함수와 가치 함수의 초기화 과정에 동시에 내재화함으로써, 보상 수렴 속도를 유의미하게 가속화하고 환경 적응성을 향상시키며, 자율 요격을 위한 해석 가능한 정책 생성을 가능하게 한다. 실험 결과는 LLM 기반 추론과 강화학습 제어기의 결합이 차세대 공중 전장 환경에 적합한 강인하고 확장 가능하며 에너지 효율적인 자율 방어 시스템을 구축하기 위한 새로운 기반을 제공할 수 있음을 시사한다.

I. 서론

21세기 이후의 전쟁 양상은 정보 중심·비접촉·자율화된 전장으로 빠르게 진화하고 있으며, 그 중심에는 무인항공기(Unmanned Aerial Vehicle, UAV), 즉 드론이 자리하고 있다. 특히 러시아-우크라이나 전쟁은 드론이 정찰 및 정밀 타격 수단으로서 전통적 무기체계를 압도할 수 있음을 보여주었으며, 저비용·고효율 구조의 드론이 전장 전반의 전략 균형을 바꾸는 비대칭 전력으로 기능함을 입증하였다 [1, 2]. 이러한 변화는 현대 방공체계가 직면한 새로운 기술적·경제적 위협의 본질을 드러낸다. 기존의 방공체계는 주로 지대공 미사일(Surface-to-Air Missile, SAM)이나 유도탄 요격 시스템(예: Iron Dome)과 같은 고가의 요격체계에 의존해 왔다. 그러나 이러한 체계는 고비용 구조와 제한된 기동 범위로 인해, 소형 드론의 대량 침입(Swarming)이나 기습적 공격에 취약하다. 특히 단가 수천만 달러의 미사일로 수백 달러 수준의 드론을 요격하는 것은 경제적 비대칭을 야기하여, 장기전 수행 능력을 근본적으로 저하시킨다 [3]. 이에 따라, 물리적 충돌 또는 네트워크 기반의 직접 요격을 수행하는 저비용·고기동형 안티드론 시스템(Anti-Drone System, ADS)이 새로운 대안으로 제시되고 있다 [4].

하지만 이러한 ADS 구현은 고전 제어기(Classical Controller)의 구조적 한계에 직면한다. 예를 들어, Proportional - Integral - Derivative (PID) 제어기나 Model Predictive Control (MPC)와 같은 페루프 제어기는 선형 근사 및 정적 환경을 가정하기 때문에, 실제 전장의 비선형·동적 환경에서는 적응성이 급격히 저하된다. 시스템 동역학적 특성이 변화할 때마다 제어기를 재설계해야 하며, 이는 실시간성·경제성 측면에서 비현실적이다. 이에 반해 강화학습(Reinforcement Learning, RL)은 환경과의 상호작용을 통해 정책을 스스로 학습함으로써, 사전 모델링 없이도 비선형 환경에서 적응 가능한 제어를 수행할 수 있다 [5]. 그러나 기존 RL 기법은 보상 수렴이 느리고(수백만 회 이상의 상호작용 필요), 학습 초기의 무작위 탐

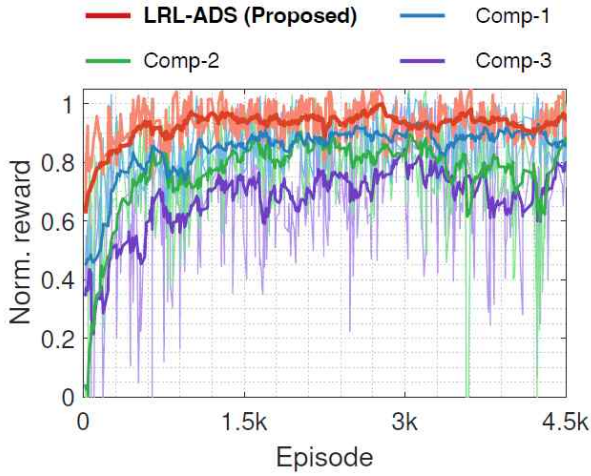
색으로 인해 정책 수렴 속도와 실시간 대응성이 제한되는 구조적 문제를 가진다. 이러한 단점은 전장의 시·공간적 변동이 빠른 실제 무인기 작전 환경에서 치명적인 제약 요인이 된다.

본 연구는 이러한 RL의 근본적 한계를 극복하기 위해, 대규모 언어 모델(Large Language Model, LLM)의 논리적 추론 및 맥락 이해 능력을 RL의 학습 구조에 결합한 새로운 제어 패러다임을 제시한다. 제안된 LLM-유도 강화학습(Large Language Model-Guided Reinforcement Learning, LRL)은 LLM이 초기 정책을 ‘지도(Guidance)’하여 학습 공간을 효율적으로 탐색하고, 가치함수(Value Function) 학습의 초기조건을 고도화함으로써 빠른 보상 수렴을 가능하게 한다. LLM은 학습 초기의 정책 학습 단계에서 정책 학습을 더욱 효율적으로 하여 더욱 강건한 초기 학습을 가능하게 한다. 일반적인 강화학습은 최적 정책을 학습하기 위해 환경과의 수많은 상호작용을 필요로 한다. 그러나 LLM-유도 강화학습은 LLM의 강건한 추론 능력을 통해 초기 정책 학습에 있어 LLM의 도움 및 안내를 받아 더욱 빠른 학습 수렴을 가능하게 한다.

II. LLM-Guided 강화학습

LLM-유도 강화학습 기반 ADS는 Partially Observable Markov Decision Process (POMDP) 구조로 모델링되며, 상태(State)는 군용 드론과 적 드론 간의 상대 거리, 잔여 에너지, 포획 여부로 구성된다. LLM은 지상 통제탑에서 초기 정책을 생성하고, RL 에이전트는 이를 바탕으로 상태 - 행동 가치함수를 학습하여 최적정책을 획득한다.

이러한 초기 정책 주입은 무작위 탐색 대신 의미 기반 탐색(Semantically Guided Exploration)을 수행하게 하여, 학습 안정성과 수렴 속도를 동시에 개선한다. 또한 에이전트는 누적 보상을 최대화하도록 설계되어 있으며, LLM은 이 보상의 그래디언트 방향을 초기화 시점에서 최적화 방향으로 유도한다



[그림 1] 알고리즘별 학습된 군용드론의 정규화된 보상값

III. 성능평가

[그림 1]은 제안된 알고리즘을 사용하여 훈련된 군용 드론의 정규화 누적 보상 추세를 벤치마크 방법과 비교하여 보여줍니다. Comp-1/2/3은 각각 Advantage Actor Critic (A2C), Reinforce, Deep Q-network (DQN)을 의미한다. 제안된 알고리즘과 벤치마크 모두에서 훈련이 진행됨에 따라 제안된 LRL-ADS로 훈련된 군용 드론이 가장 높은 정규화 보상 값을 달성합니다. 반면, 벤치마크 알고리즘으로 훈련된 군용 드론의 보상은 증가하고 결국 수렴하지만, 제안된 알고리즘으로 훈련된 드론이 달성한 보상 값에 비해 낮은 값에서 수렴합니다. RL 공식화에서 보상 함수는 군용 드론의 행동을 원하는 목표로 안내하도록 미리 설계되어 있으므로 보상 값이 높을수록 우수한 훈련 성과와 강력한 제어 능력을 나타냅니다. 수렴 후에도 제안된 알고리즘은 Comp-1, Comp-2, Comp-3보다 각각 1.11배, 1.12배, 1.24배 더 높은 0.957의 가장 높은 정규화 보상 값을 달성한다. 여기서 가장 주목할 만한 점은 제안된 알고리즘이 다른 벤치마크에 비해 더 빠른 훈련 수렴 속도를 보여준다는 것이다. 즉, [그림 1]에서 볼 수 있듯이 제안된 알고리즘은 가장 적은 에피소드 내에서 보상 수렴을 달성하여 LLM 기반 정책 훈련의 주요 장점을 강조한다. 제안된 알고리즘이 달성한 우수한 보상 값은 지상 지휘 센터의 LLM 시스템을 통해 전장 환경에서도 신속한 훈련을 가능하게 하는 견고한 ADS를 구축할 수 있음을 나타낸다. 이러한 실험적 결과는 전장환경에서 제안된 알고리즘의 강건한 안티 드론 성능을 가짐을 의미한다. LLM은 학습 초기에 RL 에이전트인 군용 드론이 효율적인 정책함수 학습을 할 수 있도록 도와준다. 이러한 LLM의 역할은 안티 드론 임무를 수행하는 군용 드론이 보다 효율적으로 행동 공간을 탐색하고 올바른 행동을 취할 수 있도록 도와준다.

iv. 결론

본 논문은 대규모 언어 모델의 추론·계획 능력을 강화학습의 정책 학습 과정에 결합한 LLM-유도 강화학습 기반 안티드론 방공 시스템을 제안하였다. 제안된 방법은 LLM이 생성한 초기 정책을 RL 학습에 통합함으로써 보상 수렴 가속을 달성하였다. 실험 결과, 기존 강화학습 기반 제어기 대비 수렴 속도와 제어 안정성이 현저히 개선되었으며, 실제 항공역학 모델 기반 시뮬레이션에서도 강건성과 적응성을 유지하였다. 따라서 본 연구는 지능형 군용 무인기 제어 및 자율 방공 시스템 설계의 핵심 기반 기술로서, 향후 다중 에이전트 협력(Multi Agent Reinforcement Learning, MARL)/LLM-기반 의사결정 프롬프트 최적화 등으로 확장될 수 있는 토대를 마련하였다.

ACKNOWLEDGMENT

본 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원-대학ICT연구센터(ITRC)의 지원을 받아 수행된 연구임(IITP-2026-RS-2024-00436887). 본 논문의 교신저자는 김중현임.

참고 문헌

- [1] G. S. Kim, Y. Cho, S. Park, S. Jung, and J. Kim, "Quantum Multiagent Reinforcement Learning for Joint Cube Satellites and High-Altitude Long-Endurance Aerial Vehicles in SAGIN," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 61, no. 4, pp. 9490-9510, August 2025.
- [2] M. Lee et al., "A Study on the Advancement of Intelligent Military Drones: Focusing on Reconnaissance Operations," *IEEE Access*, vol. 12, pp. 55964-55975, April 2024.
- [3] A. Adel, N. H. S. Alani, S. Thompson Whiteside and T. Jan, "Who is Watching Whom? Military and Civilian Drone: Vision Intelligence Investigation and Recommendations," *IEEE Access*, vol. 12, pp. 177236-177276, November 2024.
- [4] D. Chauhan, H. Kagathara, H. Mewada, S. Patel, S. Kavaiya and G. Barb, "Nation's Defense: A Comprehensive Review of Anti-Drone Systems and Strategies," *IEEE Access*, vol. 13, pp. 53476-53505, March 2025.
- [5] G. S. Kim, S. Lee, T. Woo, and S. Park, "Cooperative Reinforcement Learning for Military Drones over Large-Scale Battlefields," *IEEE Transactions on Intelligent Vehicles*, pp. 1-11, 2024 (Early Access).