

다중 에이전트 심층 강화학습을 활용한 프로토콜 학습 연구 동향

박상준

광운대학교 전자통신공학과

sangjunpark@kw.ac.kr

A Survey on Protocol Learning Using Multi-Agent Deep Reinforcement Learning

Sangjun Park

Kwangwoon University

요약

본 논문은 Miuccio 외의 연구 [1]를 소개하며, 다중 에이전트 강화학습(MARL)을 활용한 무선 MAC 프로토콜 학습 기법과 그 특징을 살펴본다. 해당 연구는 실행 가능성과 일반화 성능을 고려한 학습 프레임워크를 제안하고, 시뮬레이션을 통해 기존 방식 대비 향상된 성능을 보였다. 또한 본 논문에서는 이와 관련된 향후 연구 방향도 함께 고찰한다.

I. 서론

본논문에서는 Miuccio 외의 연구 “On Learning Generalized Wireless MAC Communication Protocols via a Feasible Multi-Agent Reinforcement Learning Framework” [1]을 토대로, 무선 MAC 프로토콜 설계를 위한 최신 학습 기반 접근 방식을 소개 및 분석하고, 이를 바탕으로 향후 연구 방향에 대해 간단히 논의한다.

II. 본론

본논문에서 소개하는 연구는 차세대 6G 무선 통신 환경에서 사용자 요구와 네트워크 상황의 복잡성을 반영할 수 있는 학습 기반 MAC 프로토콜 설계의 필요성을 강조한다. 특히 기존 연구들이 대부분 이상적인 시뮬레이션 환경을 가정하여 실제 시스템에 적용하기 어렵고, 학습된 정책이 훈련된 환경을 벗어났을 때 일반화되지 않는 문제점을 지적한다. 이러한 한계를 극복하고자, 본 논문에서는 현실적인 적용이 가능하면서도 높은 일반화 성능을 갖춘 MAC 프로토콜 학습 프레임워크를 제안한다.

제안된 접근법의 핵심은 통신 시스템을 다중 에이전트 부분 관찰 마르코프 결정 과정(Multi-agent Partially Observable Markov Decision Process; MPOMDP)으로 모델링하는 것이다. 사용자 단말(UE)은 매 시간 슬롯마다 데이터 전송(dPDU), 삭제, 대기 등의 동작을 수행할 수 있으며, 제어 평면에서는 시그니처 기반의 스케줄링 요청(sPDU)을 송신한다. 기지국(BS)은 이전 슬롯에서의 수신 결과에 따라 ACK 또는 Scheduling Grant(SG)를 전송하여 통신 자원을 분배한다.

전체 프레임워크는 중앙집중 학습- 분산 실행(Centralized Training, Decentralized Execution; CTDE) 구조를 따르는 다중 에이전트 강화학습(MARL) 방식으로 구성되며, 각 단말은 자체 관찰 정보만을

기반으로 행동을 결정한다. 정책은 Proximal Policy Optimization(PPO) 알고리즘을 통해 학습된다. 이 프레임워크는 다음과 같은 측면에서 기존 접근 방식과 차별화된다:

- 현실 적용 가능성 확보: 중앙집중적인 전역 보상 계산 방식 대신, BS 가 관측할 수 있는 정보만을 활용하여 보상 함수를 정의함으로써 실제 무선 통신 환경에서도 적용이 가능한 학습 구조를 구성하였다.
- 정책의 일반화 능력 향상: 고차원 관찰 데이터를 처리하는 대신, 관찰 추상화(observation abstraction)를 통해 필수 정보만을 추출함으로써, 사용자 수의 변화, 트래픽 부하, 채널 상태 등 다양한 조건 변화에 유연하게 대응 가능한 정책 학습을 가능하게 하였다.
- 분산 네트워크 환경에 적합한 학습 구조: 각 에이전트는 독립적인 분산형 크리티 네트워크를 유지하여, 협력 기반 문제 설정에서도 중앙화 없이 안정적인 학습이 가능하도록 설계되었다.

시뮬레이션 결과에 따르면 제안된 학습 기반 MAC 프로토콜은 기존의 고정형 방식에 비해 데이터 전송 성공률을 향상시키고 충돌 발생률을 크게 감소시키는 우수한 성능을 나타냈다. MARL 기법이 네트워크 계층 설계에 효과적으로 활용될 수 있음을 보여준다.

III. 향후 연구

본 연구는 MARL 을 활용한 MAC 프로토콜 학습의 유효성을 보여주었으나, 실환경 적용을 위해서는 추가적인 기술적 보완이 필요하다.

- 화장성 확보: 사용자 수의 증가 및 제한된 시그니처 자원 하에서도 학습 프레임워크의 성능을 안정적으로 유지하기 위해, 분산형 학습 구조 도입, 네트워크 클러스터링, 그리고 통신

- 오버헤드를 줄이기 위한 효율적인 메시지 교환 방식이 필요하다.
2. 과업 중심 학습 구조: 단순한 데이터 전송 성능이 아닌, 특정 과업의 성공률을 최적화하는 방향으로 강화학습 프레임워크를 재설계할 필요가 있다.
 3. QoS 및 안정성 보장 학습: 다양한 서비스 유형별 트래픽 특성, 우선순위, 지연 요구사항을 충족시키는 동시에, 시스템의 안정성과 예측 가능성을 확보할 수 있는 강화학습 기반 정책 설계가 요구된다.
 4. 계층 간 통합 최적화: MAC 계층을 넘어 물리 계층과의 연동을 고려한 통합 학습 접근을 통해 시스템 전체의 성능을 더욱 효과적으로 개선할 수 있다.

IV. 결론

본 논문은 무선 MAC 프로토콜 설계를 위한 MARL 기반 학습 프레임워크의 핵심 개념과 장점을 소개하였다. 제안된 방식은 실제 적용 가능한 학습 구조와 뛰어난 일반화 성능을 바탕으로, 기존 방식 대비 향상된 전송 성공률과 낮은 충돌률을 보였다. 향후에는 스케일 확장, 과업 중심 학습, QoS 보장, 계층 간 최적화 등을 통해 보다 실용적인 통신 시스템 구현이 기대된다.

ACKNOWLEDGMENT

"본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학 ICT 연구센터사업의 연구결과로 수행되었음"
(IITP-2025-RS-2023-00258639)

참 고 문 현

- [1] L. Miuccio, S. Riolo, S. Samarakoon, M. Bennis, and D. Panno, "On Learning Generalized Wireless MAC Communication Protocols via a Feasible Multi-Agent Reinforcement Learning Framework," *IEEE Transactions on Machine Learning in Communications and Networking*, vol. 2, 2024.