

강화학습 기반 상향링크 자원 할당 연구 동향 조사

유정엽, 고영채
고려대학교

youj0305@korea.ac.kr, koyc@korea.ac.kr

A Survey on Reinforcement Learning Based Uplink Resource Allocation

Jeong-Yeop You, Young-Chai Ko
Korea Univ.

요 약

5 세대 이동통신 네트워크는 기하급수적으로 증가하는 무선 데이터 트래픽과 다수의 사용자 서비스를 지원해야 하는 과제를 안고 있다. 특히 상향링크에서 전력 제어 및 자원 할당의 효율성이 전체 시스템 성능에 큰 영향을 미치며, 이를 해결하기 위해 강화학습 기반 기술이 주목받고 있다. 본 논문에서는 상향링크 자원 할당 문제에 적용된 다양한 강화학습 기반 접근법을 소개하고, 각 기법의 구조적 특징과 성능을 비교 분석하며 향후 연구 방향을 제시한다.

I. 서 론

5 세대 (5G) 이동통신 네트워크는 무선 데이터 트래픽이 기하급수적으로 증가하였고, 다양한 종류의 서비스를 사용하는 다수의 모바일 사용자를 지원해야 한다. 특히 상향링크에서 전체 시스템의 성능 및 서비스 품질 (QoS)을 최적화하기 위해 적절한 전력 및 자원 할당 알고리즘을 설계해야 한다. 기존의 일부 연구에서는 딥러닝 기반의 데이터 기반 접근법을 사용하여 계산 복잡도를 줄이고자 하였지만, 정확한 데이터셋 및 최적해를 확보하기 힘들고 학습 과정에서 시간이 많이 소요된다는 한계를 가지고 있다 [1]. 이러한 문제를 해결하기 위하여 연구되고 있는 방법 중 하나는 강화학습 기반 상향링크 자원 할당 알고리즘이다. 해당 기술은 주어진 환경에 대하여 에이전트 (Agent)가 상호작용하며 장기적인 시스템 성능을 극대화하는 정책을 학습하고, 이는 실시간 동적 자원 할당 문제를 해결할 수 있는 유망한 방법이다. 본 논문에서는 강화학습 기반 상향링크 자원 할당에 대한 연구 동향을 살펴본다.

II. 본론

강화학습 기반 상향링크 자원 할당을 연구한 다음 논문들은 무선 통신 상황에서 전력, 리소스 블록 (Resource Block)을 효율적으로 분배하여 스펙트럼 효율 (Spectral Efficiency) 및 통신 단절 확률 (Outage Probability)를 최적화하는 데 목적이 있다. [2]에서는 상향링크 NOMA (Non-Orthogonal Multiple Access) 시스템에서 송신 전력을 최소화하면서도 사용자 QoS 를 만족시키는 자원 할당 정책을 학습하는데 목적이 있다. 특히 이산적인 액션만을 처리할 수 있고, 액션의 수가 많아질수록 출력 차원이 기하급수적으로 증가한다는

단점을 가진 DQN (Deep Q-Learning)을 대체하기 위한 학습 방법으로 DDPG (Deep Deterministic Policy Gradient) 방법을 사용한다. 제안된 방법은 [그림 1]과 같이 먼저 DQN 을 활용하여 이산 변수인 서브캐리어 할당을 수행하고, 이어서 DDPG 를 이용해 연속 변수인 송신 전력 할당을 처리하는 구조를 갖는다. DQN 기반 서브캐리어 할당은 ϵ -greedy 정책을 통해 이산적인 사용자-서브캐리어 매핑을 출력하고, DDPG 기반 전력 제어는 연속적인 송신 전력을 각 사용자에게 대하여 동적으로 출력한다. 이때 액터-크리틱 (Actor-Critic) 구조를 활용하여 학습 안정성을 확보하고, 타깃 네트워크 및 리플레이 버퍼 (Replay Buffer)를 통해 샘플 간 상관성을 줄인다. 각 타임 슬롯에서 선택된 자원 할당 결과에 따라 시스템의 총 에너지 효율을 보상으로 계산하고, 이 보상을 장기 누적 성능으로 극대화하는 방향으로 정책이 학습된다. 제안된 알고리즘의 성능은 상향링크 다중 사용자 NOMA 시뮬레이션을 통해 검증되었다. 사용자는 셀 내에서 무작위로 이동하며, 최대 전력을 사용하는 고정 전력 방식이며, 평가 지표로는 수렴 속도, 평균 보상, 에너지 효율 등이 사용되었다. 시뮬레이션 결과, 제안된 알고리즘은 빠른 수렴 속도와 가장 높은 평균 에너지 효율을 달성하였고, 최대 송신 전력 변화에도 안정적으로 적응하여 우수한 성능을 유지하였다. 또한, 이산 전력 방식에 비해 액션 공간 (Action Space)의 효율적 표현과 연속성 유지, 고정 전력 방식에 비해 불필요한 에너지 소비 감소 등의 측면에서 장점을 보였다.

[3]에서는 상향링크 멀티 셀 시스템에서의 효율적 자원 RB 할당 및 송신 전력 제어를 위해 다중 에이전트 심층 강화학습 (MADRL)을 기반으로 한 최적화 기법을 제안하였다. 특히 OFDMA 기반 멀티 셀 다중 사용자 환경에서 셀 간 간섭 (ICI)을 줄이면서 평균 스펙트럼 효율을 극대화하고자 하였다. 제안된 방법은 중앙 집중 학습 및 분산 실행 구조 (CTDE)를 기반으로 하며, 각 사용자에

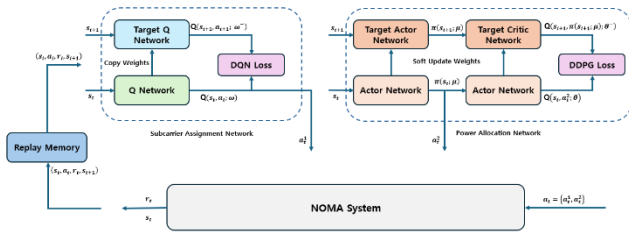


그림 1. MADRL 기반 NOMA 자원할당 시스템 모델

대해 RB 할당과 전력 제어를 담당하는 두 개의 에이전트가 존재한다. 알고리즘은 VDN (Value Decomposition Network)을 통해 각 에이전트의 Q 값을 통합하여 협력 학습 성능을 강화하고, 학습 중 에이전트의 상태에 따라 미래 보상에 대한 가중치를 동적으로 조정하는 보상 할인 네트워크 (Discount Network)를 추가로 도입하였다. 이 네트워크는 각 타임슬롯마다 보상 값을 입력으로 받아 적절한 할인 인자를 출력하여 수렴성과 안정성 향상에 기여한다. 시뮬레이션은 [그림 2]와 같이 7 개의 셀과 최대 8 명의 사용자가 있는 시스템을 가정하며, 정적 할인 기반 VDN, 독립 Q-Learning (IQL), 무작위 방식과 비교하였다. 실험 결과, 제안된 알고리즘은 평균 스펙트럼 효율에서 가장 우수한 성능을 보였으며, 사용자 수가 증가해도 강건한 수렴성과 높은 안정성을 유지하였다. 특히 할인 인자를 네트워크가 동적으로 학습함으로써, 수동 파라미터 조정 없이도 안정적인 학습이 가능함을 확인하였다.

[4]에서는 URLLC (Ultra-Reliable and Low-Latency Communication) 환경에서 상향링크와 하향링크 자원 할당을 통합적으로 최적화하기 위해 PPO (Proximal Policy Optimization) 기반의 DRL 알고리즘을 제안하였다. 제안된 시스템은 다중 사용자 MISO-OFDMA 구조에서 패킷 왕복 지연 (Round-trip delay)을 최소화하는 것을 목표로 하며, 서브채널 할당과 전송 슬롯 할당을 동시에 학습한다. 학습 구조는 상향 및 하향 각각에서 정책을 생성하는 두 개의 Actor 와 두 상태를 통합해 평가하는 공통 Critic 으로 구성된다. 또한, 학습 성능 향상을 위해 멀티헤드 네트워크 구조, 액션 마스킹, 슬롯 할당 보조 기능 (Slot Allocation Assistant), 클립 비율 적응 조절 (Adaptive Clipping Ratio) 등의 기법을 함께 도입하였다. 시뮬레이션 결과, 제안된 알고리즘은 separated PPO, DDPG, FCFS (First Come First Served) 등 기존 방법들 보다 더 짧은 상향링크 지연과 낮은 자원 낭비 비율을 달성하여 우수한 성능을 보였다.

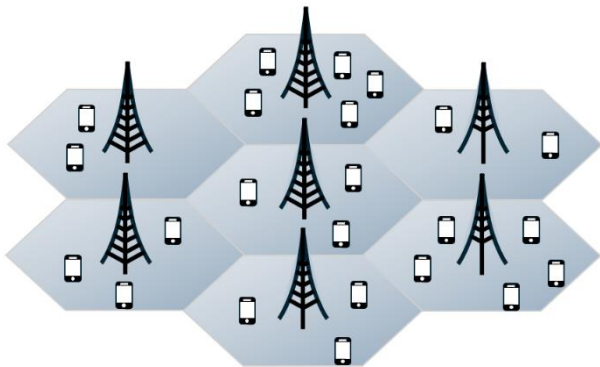


그림 2. OFDMA 기반 다중 셀 및 다중 사용자 상향 링크 셀룰러 네트워크 구성도

III. 결론

본 논문에서는 상향링크 자원 할당 문제를 해결하기 위한 강화학습 기반 기술에 대한 연구 동향을 조사하였다. 기존의 전통적인 최적화 기반 기법들은 복잡한 수학적 모델링과 높은 계산 비용, 실시간 적용의 어려움 등 여러 한계를 지니고 있으며, 특히 무선 채널의 동적 특성과 불확실성을 효과적으로 반영하지 못하는 경우가 많다. 이에 따라 최근에는 심층 강화학습을 기반으로 한 다양한 접근 방식이 제안되고 있으며, 이는 에이전트가 환경과 상호작용을 통해 자율적으로 최적의 자원 제어 정책을 학습할 수 있다는 장점을 지닌다. 본 논문에서는 DQN-DDPG 기반 이중 구조, MADRL 기반 협력 학습, PPO 기반 상향 및 하향 통합 제어 등이 대표적으로 다루어졌다. 이러한 기법들은 다양한 네트워크 조건과 사용자 수 변화에도 안정적인 수렴성과 우수한 성능을 유지하였으며, 스펙트럼 효율, 에너지 효율, 지연 최소화 등의 측면에서 기존 방식 대비 뚜렷한 성능 향상을 보였다. 이와 같이 다양한 강화학습 기반 구조를 무선 자원 할당 문제에 적용함으로써, 향후 실시간성과 확장성이 요구되는 통신 시스템에서 효과적인 자원 제어 알고리즘을 설계할 수 있을 것으로 기대된다.

ACKNOWLEDGMENT

이 논문은 2024 년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (PM2024ST0019).

참 고 문 헌

- [1] M. Liu, T. Song and G. Gui, "Deep Cognitive Perspective: Resource Allocation for NOMA-Based Heterogeneous IoT With Imperfect SIC," in *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 2885-2894, April 2019.
- [2] Y. Zhang, X. Wang and Y. Xu, "Energy-Efficient Resource Allocation in Uplink NOMA Systems with Deep Reinforcement Learning," *2019 11th International Conference on Wireless Communications and Signal Processing (WCSP)*, Xi'an, China, 2019, pp. 1-6.
- [3] Y. Yang, T. Lv, Y. Cui and P. Huang, "MADRL Based Uplink Joint Resource Block Allocation and Power Control in Multi-Cell Systems," *2023 IEEE Wireless Communications and Networking Conference (WCNC)*, Glasgow, United Kingdom, 2023, pp. 1-6.
- [4] L. Zhang, Y. Zhang and J. Zheng, "Deep Reinforcement Learning Based Joint Uplink and Downlink Resource Allocation for URLLC," in *IEEE Transactions on Vehicular Technology*, vol. 74, no. 4, pp. 6048-6063, April 2025.