

통신 음영 지역의 에너지 효율 향상을 위한 강화학습 기반 지능형 반사 표면 제어 기법 연구

김지형, 김정현*
한국전자통신연구원, *세종대학교

savant21@etri.re.kr

A Study on a RL-based RIS Control Scheme for Energy Efficiency Enhancement in Coverage Hole Environments

Kim Jihyung, Kim Junghyun*
ETRI, *Sejong Univ.

요 약

본 논문은 통신 음영 지역의 에너지 효율(EE) 최대화를 위해, 지능형 반사 표면(RIS)의 실질적인 운용 비용(전력/지연)을 보상 함수에 직접 통합하는 심층 강화학습(DRL) 기반 제어 기법을 제안한다. 제안 기법은 시뮬레이션을 통해 기존 방식 대비 낮은 SNR(신호 대 잡음비) 환경에서의 우수한 EE 성과와 개선된 복잡도-성능 효율성을 입증하였다.

I. 서 론

차세대 무선 통신 시스템의 핵심 기술 중 하나인 지능형 반사 표면(RIS)은 거의 수동적인 반사 소자를 통해 무선 채널 환경을 프로그래밍함으로써, 통신 성능을 유의미하게 향상시킬 수 있는 잠재력을 가진다 [1]. 특히, 셀 경계나 실내 음영 지역과 같이 채널 이득이 낮은 환경에서 RIS 는 비용 효율적으로 커버리지를 확장하고 신호 품질을 보장하는 핵심 조력자로 주목받고 있다.

그러나 RIS 의 실용적인 구현을 위해서는 소자의 위상 전환과 제어부 구동에 소모되는 전력 등 현실적인 운용 비용을 반드시 고려해야 한다 [2]. 신호 대 잡음비가 극도로 낮은 환경에서는 RIS 를 활성화하여 얻는 빔포밍 이득이 이러한 운용 비용을 상쇄하지 못할 수 있으며, 이 경우 RIS 를 비활성화하는 것이 시스템 전체의 EE 를 오히려 향상시키는 최적의 선택이 된다. 따라서, 실시간으로 변하는 채널 환경에 맞춰 RIS 의 성능 이득과 운용 비용 간의 복잡한 상충 관계를 실시간으로 최적화하는 동적 제어 정책이 필수적이다.

본 논문에서는 이러한 다중 목표 최적화 문제를 해결하기 위해 심층 강화학습(DRL)에 기반한 접근법을 제안한다. DRL 은 불완전한 채널 상태 정보 하에서도 동적인 자원 할당 정책을 효과적으로 학습할 수 있는 강력한 프레임워크이다. 구체적으로, 본 논문은 RIS 제어에 따른 전력 및 지연 비용을 명시적인 페널티 항으로 포함하는 새로운 보상 함수를 설계하고, 이를 기반으로 EE 를 최대화하는 TD3 에이전트를 제안한다. 제안하는 에이전트는 채널 품질에 따라 RIS 를 선별적으로 활성화하는 정책을 학습하여, 열악한 통신 환경에서 시스템의 에너지 효율을 최적화한다.

II. 시스템 모델 및 전력-지연 인지 TD3 (PL-TD3)

본 논문에서는 RIS 가 적용된 일반적인 무선 통신 환경을 고려한다. 대표적인 시스템 모델로서, 하나의 기지국과 RIS 의 도움을 받아 통신하는 단일 사용자가 존재하는 단일 셀 하향링크 환경을 가정한다. 여기서 사용자는 셀 경계나 장애물로 인해 채널 이득이 낮은 통신 음영 지역에 위치하는 상황을 상정한다. 따라서 본 논문의 목표는, RIS 운용 비용을 포함한 시스템의 총 EE 를 최대화하는 최적의 RIS 제어 정책을 찾는 것이다. 이를 위해 EE 는 식 (1)과 같이 정의된다.

$$EE = \frac{R(W, \Phi)}{P_{tx} + P_{ctrl}(\Phi) + P_{circ}(\Phi)} \quad (1)$$

여기서 $R(W, \Phi)$ 는 기지국 프리코더 W 와 RIS 위상 행렬 Φ 에 의해 결정되는 채널 용량 기반의 데이터 전송률이다. P_{tx} 는 기지국의 송신 전력이며, 본 논문의 핵심 고려사항인 $P_{ctrl}(\Phi)$ 와 $P_{circ}(\Phi)$ 는 각각 RIS 의 위상 구성 및 제어 신호 처리에 소모되는 제어 전력과 소자 및 제어부 구동에 필요한 회로 전력을 의미한다.

이러한 동적 환경에서 EE 를 최적화하기 위해, 본 논문은 PL-TD3 (Power and Latency Aware TD3) 프레임워크를 제안한다. PL-TD3 의 목표는 최적의 정책 $\pi(s_t)$ 를 학습하는 것이다. 여기서 최적의 정책이란, 각 상태에서 장기적으로 기대되는 누적 보상의 합을 최대화하는 일련의 행동을 결정하는 규칙, 즉 상태를 행동으로 매핑하는 함수를 의미한다. 에이전트의 상태 s_t 는 이전 시간대의 채널 품질 정보(CQI), 데이터 전송률, 그리고 선택했던 RIS 구성을 포함하여 현재 채널 환경을 포괄적으로 인지할 수 있도록 구성된다. 이를 바탕으로 에이전트는 RIS 의 동적 활성화를 위한 게이트 $g_t \in \{0, 1\}$ 와 코드북 $\{\Phi_m\}$ 내에서 최적의 위상 행렬을 선택하기 위한 인덱스 m_t 의 조합으로 정의되는 행동 a_t 을 결정한다. 제안 기법의 핵심인 보상 r_t 는 에이전트가

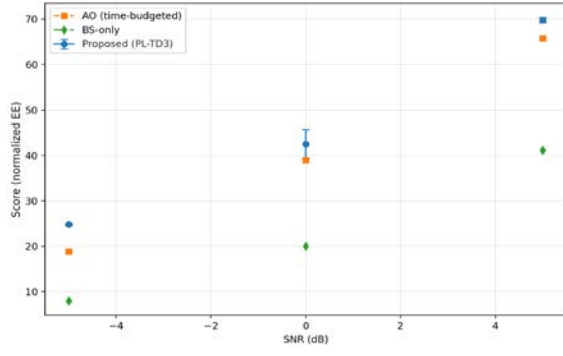


그림 1. SNR 에 따른 에너지 효율(EE)

EE 와 운용 비용 간의 상충 관계를 학습하도록 아래 식 (2)와 같이 설계된다.

$$r_t = \alpha EE_t - \lambda P_{ctrl}(\Phi_t) - \lambda_c P_{circ}(\Phi_t) - \mu \Delta t_t \quad (2)$$

(2)에서 첫 번째 항은 시스템의 주 목표인 에너지 효율이며, 이후의 항들은 각각 제어 전력, 회로 전력, 그리고 정책 결정 지연 시간에 대한 페널티를 나타낸다. 제안하는 PL-TD3 에이전트의 학습 과정은 매 시간 슬롯마다 상태 s_t 를 관측하고, 정책 $\pi(s_t)$ 에 따라 행동 a_t 를 선택하여 시스템에 적용한다. 그 결과로 관측된 성능 지표들을 바탕으로 (2)에 따라 보상 r_t 를 계산하고, 상태 천이 정보 (s_t, a_t, r_t, s_{t+1})를 리플레이 버퍼에 저장하여 액터-크리틱 신경망을 TD3 알고리즘에 따라 업데이트한다. 이러한 학습 절차는 5G/6G 이동통신 시스템의 제어 프레임워크 내에서 적용될 수 있으며, 에이전트의 상태 관측은 기지국이 기준 신호를 기반으로 채널을 측정하고 사용자로부터 CQI를 피드백받는 과정에 해당하며, 행동 적용은 결정된 RIS 구성을 기지국이 하향링크 제어 채널 등을 통해 RIS 제어기에 전달하는 과정으로 실행될 수 있다.

III. 시뮬레이션 및 결과 분석

제안하는 기법의 성능을 정량적으로 검증하기 위해, 통신 음영 지역 시나리오를 대상으로 시뮬레이션을 수행하였다. 이번 시뮬레이션의 핵심은 제안하는 에이전트가 RIS의 채널 이득과 운용 비용 간의 상충 관계를 효과적으로 학습하는지를 확인하는 것이다. 성능 지표로는 시스템의 총 EE와 정책 결정 시간을 분석하였으며, 기존 기법들과의 비교를 통해 제안 기법의 효용성을 확인하였다. 성능 비교를 위해 전통적인 반복 최적화 방식인 교대 최적화(AO)와 RIS를 사용하지 않는 BS-only를 비교 대상으로 설정하였다 [3]. 공정한 비교를 위해 AO는 제안 기법과 유사한 시간 예산 내에서 수행되도록 하였으며, BS-only는 RIS 적용에 따른 순수 성능 향상을 확인하기 위한 기준선으로 사용된다. 시뮬레이션 결과는 그림 1과 2에 나타내었다. 그림 1에서 제안하는 PL-TD3은, 특히 신호 대 잡음비가 낮은 환경에서 비교 대상들보다 우수한 EE를 달성함을 확인하였다. 이는 에이전트가 RIS 운용 비용을 보상 함수를 통해 인지하고, 채널 이득이 불충분할 경우 불필요한 활성화를 회피하도록 학습했기 때문이다. 비용을 인지하지 못하는 AO는 RIS를 기계적으로 활성화하여 오히려 EE가 저하되는 것과 대조적인 결과이다. 또한 그림 2는 제안 기법이 AO 대비 훨씬 짧은 결정 시간 내에 더 높은 EE를 달성함을 보여준다. 이는 단일 순방향 연산만으로 동작하는 DRL 추론 방식

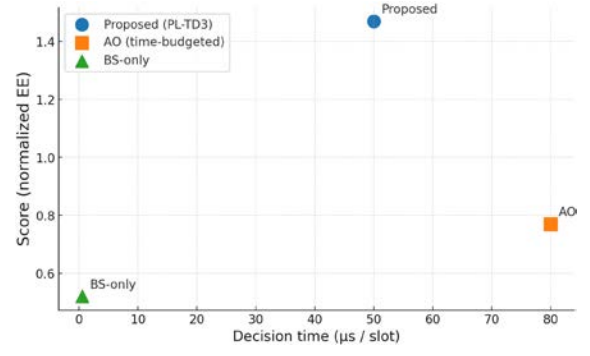


그림 2. 결정 시간과 에너지 효율(EE)

이 반복적인 최적화가 필요한 AO에 비해 실시간성에 매우 유리하기 때문이다. 결론적으로, 제안 기법은 더 나은 성능을 더 효율적으로 달성하며 실제 통신 환경에서 높은 적용 가능성을 보여준다.

IV. 결론

본 논문은 통신 음영 지역의 에너지 효율 최적화를 위해, RIS의 실질적인 운용 비용을 고려하는 심층 강화학습 기반의 PL-TD3 제어 기법을 제안하였다. 시뮬레이션 결과, 제안하는 기법은 기존 교대 최적화 방식 대비 낮은 SNR 영역에서 우수한 에너지 효율 성능을 달성함을 확인하였다. 특히 RIS를 선별적으로 비활성화하는 정책을 학습하여 불필요한 전력 소모를 효과적으로 방지하였다. 또한, 제안 기법은 AO 방식보다 훨씬 낮은 결정 시간 내에 동작하여 복잡도-성능 측면에서도 우수한 효율을 보여주었다. 이러한 결과는 제안 기법이 성능과 실시간성을 동시에 만족함을 보여주고 있다. 따라서 제안하는 PL-TD3 기법은 동적 제어가 필수적인 차세대 무선 통신 시스템의 에너지 효율을 향상시키는 효과적인 해결책이 될 수 있을 것으로 기대된다.

ACKNOWLEDGMENT

본 논문은 2025년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No. RS-2022-II220436, 지능형 재구성 표면 중계기 표준기술개발)

참고 문헌

- [1] M. Di Renzo et al., "Smart Radio Environments Empowered by Reconfigurable Intelligent Surfaces: How It Works, State of Research, and The Road Ahead," IEEE Journal on Selected Areas in Communications, vol. 38, no. 11, pp. 2450–2525, Nov. 2020.
- [2] C. Huang et al., "Reconfigurable Intelligent Surface Assisted Communications: A Power Consumption Model," in Proc. IEEE Global Communications Conference (GLOBECOM), pp. 1–6, Dec. 2020.
- [3] Q. Wu and R. Zhang, "Intelligent Reflecting Surface Enhanced Wireless Network via Joint Active and Passive Beamforming," IEEE Transactions on Wireless Communications, vol. 18, no. 11, pp. 5394–5409, Nov. 2019.