

트래픽 예측 기반 안전 심층 강화학습을 활용한 소형셀 에너지 및 QoS 다목적 최적화 기법

조아림, 윤재영, 이호원

아주대학교

{jal9160, dbswo0617, howon}@ajou.ac.kr

Multi-Objective Optimization Techniques for Small Cell Energy and QoS Using Traffic Prediction-Based Safe Deep Reinforcement Learning

Arim Cho, Jaeyoung Yun, Howon Lee

Ajou Univ.

요약

본 논문은 안전 심층 강화학습에 기반한 다단계 슬립 모드 제어 기법을 제안한다. 제안 방안은 long short-term memory (LSTM) 기반 트래픽 예측을 통합하여 소형셀 기지국의 불필요한 활성화를 줄이고, 장기적인 슬립으로 인한 quality-of-service (QoS) 저하를 사전에 완화하는 것을 목표로 한다. 이를 위해 허용되지 않는 모드 전환을 제한하는 안전 제약을 설계하고, 에너지-QoS 간 trade-off를 반영한 보상 구조를 통해 정책을 학습한다. 시뮬레이션 결과, 제안 방안은 동적 트래픽 환경에서 벤치마크 대비 에너지 소비를 줄이면서, 동시에 QoS 저하를 효과적으로 완화함을 검증한다.

I. 서론

모바일 데이터 사용량의 급증으로 무선 네트워크 트래픽이 폭발적으로 증가하며, 이에 따른 경제적·환경적 부담이 커지고 있다. 특히 기지국은 전체 네트워크 에너지 소비의 약 60~80%를 차지하므로, 보다 효율적인 제어가 요구된다 [1],[2]. 그러나 다수의 소형셀 기지국은 최대 트래픽을 기준으로 설계되어 낮은 트래픽 부하의 시간대에도 활성 상태를 유지함으로써 불필요한 에너지 낭비를 초래한다 [3]. 이러한 문제를 완화하기 위해 낮은 트래픽 부하를 갖는 기간에는 슬립 모드로 전환하는 등 에너지 효율적인 동적 운영 전략이 주목받고 있다 [4]. 더불어 기지국 기능을 중앙유닛 (CU)과 분산유닛 (DU)으로 분해한 virtualized RAN (vRAN) 아키텍처는 트래픽 조건에 따라 유연한 자원 할당을 가능하게 하여 RAN의 운영 효율성을 한층 향상시킨다 [5].

II. 안전 심층 강화학습을 활용한 소형셀 에너지 및 QoS 최적화 기법

본 논문에서는 vRAN 환경에서 시변적인 트래픽 조건에서도 에너지 소비를 줄이면서 QoS를 유지할 수 있는 기지국 슬립 모드 제어 기법을 제안한다. Long short-term memory (LSTM) 기반으로 트래픽을 예측하여 가까운 미래의 트래픽 변화 예측을 동시에 반영함으로써 동적 네트워크 환경에서 안정적이고 적응적인 슬립모드 제어가 가능하다. 지속적으로 높은 전력 소비나 과도한 슬립 상태로 인한 quality-of-service (QoS) 저하를 방지하기 위해 에너지와 QoS 간의 trade-off 관계를 명시적으로 반영한 보상 구조를 도입하며, 허용되지 않는 모드 전환을 방지하는 제약 조건을 포함한 안전 심층 강화학습 프레임워크를 도입하여 안정적인 다목적 최적화를 수행한다. 이러한 문제를 deep Q-network (DQN)로 풀기 위해 Markov decision process (MDP)로 정의한다.

- Agent: 네트워크에 배치된 모든 소형셀, $SBS(n) \in \{1, \dots, S\}$.
- State: 이전 타임스텝의 모드 전환 중 소비된 전력과 전환 시간, 선택된 모드 전환이 완료되기까지 남은 타임스텝 수, 이전 타임스텝의 트래픽, $2\tau+1$ 길이의 예측된 트래픽 정보로 구성된다.

$$s_n^t = \{P_n^{t-1}, T_n^{t-1}, NT_n^t, \rho_n^{t-1}, \rho_n^{t-\tau, t+\tau}\}.$$

- Action: $a_n^t = \{\text{active}, \text{microsleep}, \text{lightsleep}, \text{deepsleep}\}$.
- Reward: 에너지와 QoS 간의 trade-off 관계를 고려하기 위해 w_1, w_2, w_3 의 가중치를 도입한다. 보상은 에너지 절감에 따른 보상 (r_{ps}), 모드 별 전환 시간에 대한 패널티 (r_{td})와 QoS 관련 서비스 보상 (r_{sr})으로 구성되며, 전체 보상은 다음과 같이 정의된다.

$$r_n^t = w_1 \times r_{ps} + w_2 \times r_{td} + w_3 \times r_{sr}, \quad x \in \text{action set}. \quad (1)$$

$$r_{ps} = \sum_{n=1}^S \frac{P_{active}^{\max} - P_{n,x}}{P_{active}^{\max}}, \quad r_{td} = \sum_{n=1}^S \frac{T_{deep} - T_{n,x}}{T_{deep}}, \quad r_{sr} = \sum_{n=1}^S \frac{\rho_{n,x}}{\rho_n^t}.$$

III. 시뮬레이션 결과 및 결론

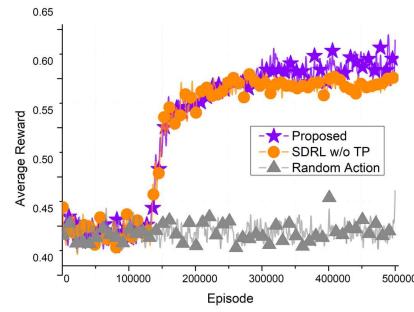


그림 1. Average reward vs. episode.

본 논문에서는 트래픽 예측을 활용한 안전 강화학습 기반 슬립모드 제어 기법을 제안하였다. 시뮬레이션 결과, 제안 방안이 트래픽 예측을 활용하지 않는 방안 (SDRL w/o TP)과 랜덤한 액션을 수행하는 방안 (Random Action) 대비 높은 값의 보상으로 수렴하는 것을 통해 에너지와 QoS 간의 trade-off를 고려한 다목적 최적화가 수행됨을 확인하였다.

ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원 (No.RS-2025-00563401, 3차원 공간에서 에너지 효율적 멀티레벨 AI-RAN 구현을 위한 AI-for/and-RAN 핵심 원천기술 연구), 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원 (No.RS-2025-02303435), 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원 (No.RS-2024-00396992, 저궤도 위성통신 핵심 기술 기반 큐브 위성 개발)을 받아 수행된 연구임.

참고 문헌

- [1] H. Lee et al., "Towards 6G hyper-connectivity: Vision, challenges, and key enabling technologies," in *Journal of Communications and Networks*, vol. 25, no. 3, pp. 344–354, June 2023.
- [2] J. Liu, B. Krishnamachari, S. Zhou, and Z. Niu, "Deepnap: Data-driven base station sleeping operations through deep reinforcement learning," *IEEE Internet of Things Journal*, vol. 5, no. 6, pp. 4273–4282, Dec. 2018.
- [3] Q. Wu, X. Chen, Z. Zhou, L. Chen, and J. Zhang, "Deep reinforcement learning with spatio-temporal traffic forecasting for data-driven base station sleep control," *IEEE/ACM Transactions on Networking*, vol. 29, no. 2, pp. 935–948, Apr. 2021.
- [4] C. Liu, B. Natarajan and H. Xia, "Small Cell Base Station Sleep Strategies for Energy Efficiency," in *IEEE Transactions on Vehicular Technology*, vol. 65, no. 3, pp. 1652–1661, March 2016.
- [5] Samsung, "White paper: vRAN value proposition and cost modeling," pp. 1–2, Sep. 2020.