

## 트래픽 예측을 활용한 다중 에이전트 강화학습 기반 기지국 전력 제어 기법

김지희, 윤희문, 이호원

아주대학교 전자공학과

jihee5034@ajou.ac.kr, gmlans1202@ajou.ac.kr, howon@ajou.ac.kr

Base Station Power Control Based on Multi-Agent Reinforcement Learning  
Based on Traffic Prediction

Ji Hee Kim, Hui Mun Yoon, Howon Lee

Ajou Univ.

## 요약

5G 및 6G로 진화한 이동통신 환경에서 트래픽의 증가로 소형 기지국이 고밀도로 배치되었지만, 최대 부하를 기준으로 설계된 네트워크는 비혼잡 시간대에서도 동일한 전력을 소비해 상당한 에너지 낭비가 발생하고 있다 [1]. 본 논문은 이러한 비효율성을 완화하기 위해, 실제 트래픽 시계열을 Transformer 모델로 예측하고 그 예측 결과를 상태(state)에 반영하여 기지국 송신 전력을 제어하는 MAQL(Multi-Agent Q-Learning) 기반 알고리즘을 제시한다. 시간대별 트래픽 부하에 따라 기지국 전력을 동적으로 조정함으로써 에너지 효율을 향상시키고, 동시에 서비스 품질(QoS)을 보장하는 것을 그 목표로 한다.

## I. 서론

본 논문에서는 Telecom Italia 실측 데이터를 기반으로 시공간 상관성을 학습하는 Transformer 기반 트래픽 예측 모델을 설계하고 이를 Multi-Agent Q-Learning(MAQL) 전력 제어 기법과 연계하는 분산형 절전 프레임워크를 제안한다. 예측-제어 통합 구조 제안을 통해 중앙집중식 스케줄링 없이 독립 학습과 결정으로 스케줄링 부하를 줄이면서도 실시간 전력 최적화를 달성하고자 한다.

## II. 시스템 모델

## A. Transformer 기반 트래픽 예측

원본 트래픽 로그는 10분 간격으로 기록된 시간 정보와 100×100 공간 그리드(grid)로 구성되어 각 셀의 실사용량(Mbps)을 제공한다. 예측 모델은 이 중 10×10 하위 영역을 추출하여 분석을 진행하였으며, 각 그리드는 하루 144개 타임스텝으로 이루어진 시계열 데이터를 가진다. 논문에서 상용한 LSTM(Long Short-Term Memory) 모델 [2]은 각 그리드의 시계열을 독립적으로 처리하여 인접 셀 간 공간 상관성을 반영하기 어렵다는 한계가 있다. 반면, Transformer 모델 [3]은 시공간 어텐션 메커니즘을 통해 시간 축뿐 아니라 공간 축에서도 상호 관계를 효과적으로 학습할 수 있다.

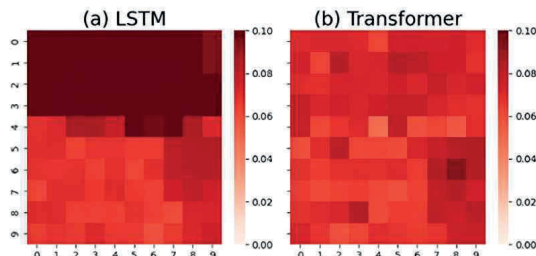


Fig 1. LSTM과 Transformer의 grid-wise 예측 오차(MAE)

예측 성능을 비교한 결과, Transformer 모델은 대부분의 그리드에서 LSTM 대비 낮은 평균 절대 오차(MAE)를 기록하였다. Fig. 1의 히트맵은 두 모델의 MAE를 그리드별로 시각화한 것으로, Transformer의 오류 전반이 LSTM보다 옅은 색조로 나타나 우수한 정확도를 직관적으로 보여 준다. 따라서 Transformer가 시공간적 상관성을 효과적으로 학습에 반영한다는 사실을 입증할 수 있다.

## B. 다중 에이전트 큐러닝 기반 전력 제어 기법

Transformer가 예측한 시공간 트래픽을 MDP(Markov decision process)의 상태(state)로 사용해 각 소형 기지국(SBS)이 송신 전력을 자율 조정하는 다중 에이전트 강화학습 기반 분산형 절전 프레임워크를 제안한다. 강화학습을 통해 예측된 트래픽 패턴에 적응하는 전력 제어 정책을 학습하기 위해 본 논문에서는 다음과 같이 Markov decision process(MDP)를 정의한다 [4].

- Agent: N개의 SBS(small cell base station)로 정의된다.

초고밀도 스몰셀 네트워크(UDSCN)에서 제안된 다중 에이전트 강화학습 프레임워크는 각 SBS를 하나의 에이전트로 간주한다. 각 에이전트는 Transformer가 예측한 슬롯별 트래픽 부하를 상태(state)로 받아 독립적으로 송신 전력을 조정하며, MBS는 간섭원으로 간주된다.

- State

각 에이전트는 자신의 상태 정보를 다른 에이전트와 공유하지 않고 오직 자신의 송신 전력과 예측 트래픽 부하만을 고려한다.

$$S = (\rho_{t,N}^{bin} - 1) \times \left( \frac{P_{max} - P_{min}}{\Delta P_t} + 1 \right) + i_{t,N}, \rho_{t,N}^{bin} \in \{1, 2, 3, 4\} \quad (1)$$

$$i_{t,N} = \frac{P_{t,N} - P_{min}}{\Delta P_t} + 1 \quad (2)$$

현재 전력  $p_{t,N}$ 을  $\Delta P_t$  간격 이산 집합  $\{P_{min} : \Delta P_t : P_{max}\}$ 에 매핑(mapping)하여 전력 인덱스  $i_{t,N}$ 을 얻고 Transformer가 예측한  $H$  슬롯 평균 부하를 이산화하여  $\rho_{t,N}^{bin}$ 을 구한다. 계산한  $\rho_{t,N}^{bin}$ 을 상위 자릿수,  $i_{t,N}$ 을 하위 자릿수로 부호화하여 상태 공간을 구성한다.

- Action

각 에이전트는 한 슬롯마다 다음의 세 가지 동작 집합  $A$  중 하나를 선택할 수 있다. 전력 조정 동작은 송신 전력 증가  $\Delta P_t$ , 송신 전력 감소  $-\Delta P_t$ , 그리고 현재 송신 전력 유지  $\Delta_0$ 로 다음과 같이 정의된다.

$$A = \{-\Delta P_t, \Delta_0, \Delta P_t\} \quad (3)$$

- Reward

전체 네트워크의 에너지 효율  $\xi(n)$ 을 극대화하기 위해 각 에이전트는 자신의 보상 정보를 다른 에이전트와 공유한다고 가정한다. 설계한 Outage-aware(인식) 보상 함수는 슬롯  $t$ 에서 전체 사용자 수  $|U|$ 와

## 참 고 문 헌

- [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] Q. Wu, X. Chen, Z. Zhou, L. Chen and I. Zhang, "Deep Reinforcement Learning With Spatio-Temporal Traffic Forecasting for Data-Driven Base Station Sleep Control," in *IEEE/ACM Transactions on Networking*, vol. 29, no. 2, pp. 935-948, April 2021.
- [3] 김태정, 나지현, (2024-06-19). 이동통신 기지국의 리소스 관리를 위한 인공지능 모델 연구. 한국통신학회 학술대회논문집, 제주.
- [4] E. Kim, H. Choi, H. Kim, I. Na and H. Lee, "Optimal Resource Allocation Considering Non-Uniform Spatial Traffic Distribution in Ultra-Dense Networks: A Multi-Agent Reinforcement Learning Approach," in *IEEE Access*, vol. 10, pp. 20455-20464, 2022.

Outage 사용자 수  $U_{out}$ 에 기반한 지수형 패널티(penalty)와 각 스몰셀 기지국의 에너지 효율로 구성된다. 에이전트  $j$ 의 보상  $R_j^c$ 는 다음과 같이 정의된다.

$$R_j^c = e^{-\frac{U_{out}}{|U|}} \times \sum_{n \in N} \xi(n) \quad (4)$$

## III. 시뮬레이션 결과

본 논문에서는 초고밀도 스몰셀 네트워크(UDSCN)를 가정하여 3개의 대형 기지국(MBS), 6개의 소형 기지국(SBS), 총 36명의 사용자가  $10 \times 10$  격자 상에 분포한 환경을 고려한다. 각 에피소드는 3,000개의 시간 슬롯으로 구성되며, 학습률 및 탐험률은 에피소드 진행에 따라 지수적으로 감소하도록 설계되었다. 제안한 predictive MAQL 기법의 성능은 트래픽 정보 없이 학습하는 no-traffic, 무작위 선택 기반의 random, 이진 on-off 제어 알고리즘과 성능을 비교 평가하였다.

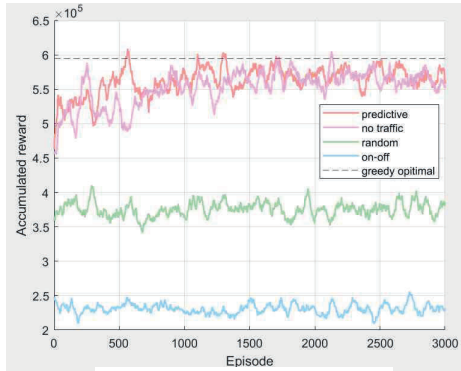


Fig 2. 에피소드별 누적 평균 보상

Transformer 예측을 반영한 predictive MAQL 곡선은 초기 에피소드에서 급격히 상승하고 준최적(greedy) 기준선 아래에서 안정적으로 수렴하는 추세를 보인다. 트래픽 정보를 제외한 no-traffic 곡선도 상승 흐름을 보이지만, 수렴 구간이 한 단계 낮아 트래픽 수요 인식이 학습 효율 향상에 기여함을 보여준다. 반면 random과 on-off 방안은 학습 전 구간에서 거의 평평하게 유지되어 휴리스틱과 이진 제어 기법의 학습 한계를 보여준다. 따라서 예측 부하를 반영한 predictive MAQL이 학습 속도와 에너지 효율 모두에서 우수한 성능을 보임을 확인할 수 있다.

## IV. 결론

차세대 통신 환경에서는 트래픽의 규모와 복잡성이 한층 가증될 것으로 예상하며, 이에 따라 지능형 기지국 제어 기술의 중요성도 더욱 부각될 것이다. 본 논문에서는 예측-제어 통합 모델이 6G 네트워크 환경에서 자원 효율화와 에너지 절감을 달성할 수 있는 유망한 접근임을 확인하였다. 향후 연구에서는 예측 기반 강화학습의 효과를 극대화하기 위한 환경 구성을 심화할 계획이다.

## ACKNOWLEDGMENT

이 논문은 2024년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원(No. RS-2024-00396992, 저궤도 위성통신 핵심 기술 기반 큐브위성 개발)과 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원(No. 2022-0-00704, 초공간 이동체 지위를 위한 3D-NET 핵심 기술 개발)과 2025년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원(RS-2025-00563401, 3차원 공간에서 에너지 효율적 멀티레벨 AI-RAN 구현을 위한 AI-for/and-RAN 핵심 원천기술 연구)을 받아 수행된 연구임.