

강화학습을 이용한 동적 스펙트럼 할당에 관한 연구

노원우, 샤흐란라이한무함마드, 최계원*
성균관대학교

daoproject@skku.edu, raihansya13@g.skku.edu *kaewonchoi@skku.edu

A Study on the Dynamic Spectrum Allocation using Reinforcement Learning

Ro Won Woo, Raihan Muhammad Syahran, Choi Kae Won*
Sungkyunkwan Univ

요약

본 논문에서는 Proximal Policy Optimization(PPO) 알고리즘을 이용한 다중 통신 환경에서의 동적 스펙트럼 할당 알고리즘을 제안한다. 다중 통신 환경에서는 여러 개체가 동시에 데이터를 전송하고 수신하기 때문에 통신 간 충돌을 방지하기 위해 효율적으로 주파수를 할당하는 방법이 필요하다. 이를 위해 복잡한 다중 통신 환경을 그래프 형식의 데이터로 구현하고 그래프 데이터에 특화된 신경망 구조인 Graph Neural Network(GNN)를 이용하여 동적 스펙트럼 할당 알고리즘을 구성하였다. 이를 통해 복잡한 통신 환경에서 안정적인 통신 품질을 유지할 수 있음을 확인하였다.

I. 서론

현대 사회에서는 통신 기술의 발전으로 통신 환경이 급속하게 변화하고 있다. 전파 자원은 공공재적 성격의 국가 자원으로써 사용할 수 있는 주파수 대역은 한정적이다. 하지만 스마트폰, 태블릿, 센서 네트워크 등 다양한 객체들이 네트워크에 연결되어 실시간으로 데이터를 교환하고 정보를 공유하는 등 통신 객체들이 크게 늘었다. 이러한 추세로 인해 통신 환경은 이전보다 훨씬 더 복잡해졌다. 통신 환경이 복잡해질수록 휴리스틱 방법론으로는 최적의 해를 찾기 어려울 수 있다. 스펙트럼 할당 문제의 경우 한정된 자원으로 최적의 해를 찾아야 하기 때문에 강화학습을 적용한 연구가 존재한다[1]. 본 논문에서는 대규모 다중 통신 환경을 다루기 위해 그래프 형태의 데이터로 구현하였고 이를 해결하는 강화학습 알고리즘을 제안하고자 한다.

II. 본론

2.1 다중 통신 환경 구성

연구를 위한 가상의 다중 통신 환경을 구성하기 위해 통신소를 노드, 통신 링크를 엣지로 설정하여 그래프를 구성하였다. 각 통신소는 1~3개의 통신소와 연결되어 있고, 10개의 내외의 기반 통신소와 해당 통신소로부터 일정 거리 안에 흩어져있도록 그래프를 설계하였다.

해당 설정에서 통신 링크의 CIR을 측정하기 위해서는 각 통신 간의 간섭 정도를 계산해야 한다. 따라서 기존 그래프의 엣지인 통신 링크를 노드로 하고 통신 링크 간의 간섭을 엣지로 하는 간섭 그래프를 구성하였다.

동적 스펙트럼 할당을 위해서는 각 통신 링크에 주파수를 배정 해야 하기 때문에 통신 링크가 노드인 간섭 그래프를 데이터로 사용하였다.

2.2 GNN을 이용한 신경망 구조

Graph Neural Network(GNN)란 그래프 데이터를 처리하는 뉴럴 네트워크의 한 유형이다[2]. 그래프 데이터란 노드와 노드를 잇는 엣지로 이루어진 데이터 구조이며 노드 간의 연결성을 효과적으로 모델링을 할 수 있다. GNN은 주로 노드 간 연결 관계와 이웃들의 상태를 이용하여 노드나 엣지 정보를 업데이트하고 마지막 상태를 통해 결과를 추출 해낸다. 이를 위해 Message Passing 기법이 사용된다. Message Passing 기법은 각 노드마다 연결된 이웃의 상태를 전달하기 위해 사용되는 Message 함수, 전달받은 Message 함수를 집계하는 Aggregate 함수, 각 노드에서 현재 노드 상태와 Aggregate 함수로 집계된 Message를 이용하여 상태를 업데이트하는 Update 함수로 이루어져 있다.

Graph Transformer Convolution은 그래프 데이터에 대한 변환을 수행하는 Graph Convolution 연산 중 하나이다[3]. 기존 Graph Convolution 연산에 self-attention 메커니즘을 도입해 각 노드가 이웃 노드의 상태를 얼마나 반영해야 하는지를 학습한다. 이를 통해 그래프 구조를 더 효과적으로 학습할 수 있다.

2.3 MDP 정의

1) 상태 정보(state) : 그래프 정보를 상태 정보로써 사용하였다. 그래프 정보는 노드 정보인 각 노드의 주파수 할당 현황, 노드의 통신 파워와 엣지 정보인 통신 간 간섭 파워를 사용한다.

2) 행동(action) : 어느 노드에 어느 주파수를 배정할지를 결정하는 것을 행동으로 설정하였다. Actor에선 각 노드의 주파수마다 결정될 확률이 출력으로 나오고 이를 샘플링하여 행동을 결정한다.

3) 보상(reward) : 각 통신은 CIR이 22 이상인 경우 성공으로 설정하였고, 이전 행동에 비해 성공 노드 개수의 증가량을 보상으로 설정했다.

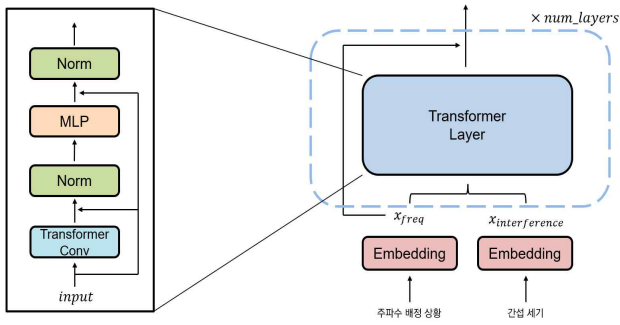


그림 1. Graph Transformer 구조

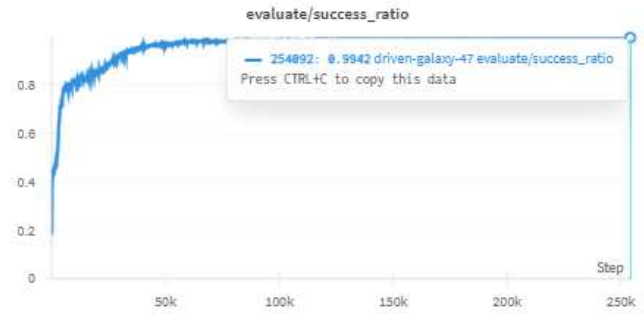


그림 2. 학습 과정 중 통신 성공 비율

2.4 GNN을 적용한 PPO 강화학습 알고리즘

PPO[4]는 Actor-Critic 구조 기반 강화학습 알고리즘으로 에이전트의 정책을 학습하는 네트워크인 Actor와 에이전트의 정책을 평가하는 네트워크인 Critic로 이루어져 있다. Actor는 현재 상태를 입력으로 받은 뒤 행동을 출력하며 Critic은 현재 상태를 입력으로 받아 상태 가치(state value)를 출력하며 이는 Actor가 좋은 행동을 선택하는 데에 사용된다. 또한 안정적인 정책 업데이트를 위해 클리핑 기법을 통해 이전 정책과 새로운 정책 사이의 차이를 제한하여 안정적인 학습을 진행한다.

본 논문에서는 그래프 데이터를 사용하기 위해 Actor와 Critic을 Graph Transformer Convolution을 사용하여 설계하였다. 그림 1은 본 논문에서 구성한 Graph Transformer 구조이다. Transformer Layer를 Graph Transformer Convolution과 MLP를 결합하여 구현하였고 이를 num_layer 개 쌓아 신경망을 구성하였다. Actor는 행동 별 확률을 출력하므로 마지막에 Linear Layer를 연결하였고, Critic은 상태 가치만을 출력하므로 mean pooling을 연결하였다.

III. 실험

통신 조건의 경우 주파수 대역은 7.3GHz, 송신 출력은 1W, 지향성 안테나 패턴 약 35dBi, 5개 파수 사용, 경로 감쇄 모델은 FSPL의 조건에서 실험하였다. 해당 조건을 통해 통신 링크의 Gain과 간섭 정도를 계산하여 간섭 그래프를 구성하여 그래프 데이터로 활용하였다.

다중 통신 환경 문제를 풀기 위해 랜덤 생성한 2500개의 그래프 데이터셋을 기반으로 PPO 모델을 학습시켰다. Actor와 Critic 모두 6 개의 Transformer Layer를 쌓았으며 공유 신경망 없이 분리하여 신경망을 구성하였다. 평가 요소로는 전체 노드 중 CIR을 22 이상을 달성한 노드의 비율로 설정하였다.

모델 학습 후 실험한 결과는 그림 2와 같다. 전체 노드 중 CIR을 22 이상 달성한 노드의 비율이 99.42%를 달성하여 다중 통신 환경에서의 동적 스펙트럼 할당 알고리즘이 성공적으로 작동함을 알 수 있다.

IV. 결론

본 논문에서는 대규모 다중 통신 환경에서의 동적 스펙트럼 할당 문제를 그래프 데이터로 변환하여 해결한 강화학습 알고리즘을 제안하였다. 이를 통해 복잡한 구조의 통신 환경에서도 간섭을 최소화하며 통신의 안정성을 확보할 수 있음을 확인할 수 있었다.

ACKNOWLEDGMENT

This work was partly supported by ICT Creative Consilience Program through the Institute of Information & Communications Technology Planning & Evaluation(IITP) grant funded by the Korea government(MSIT) (IITP-2024-2020-0-01821, 50%) and This work was supported by the BK21 FOUR Project.(50%)

참고 문헌

- [1] Ding, R., Gao, F., & Shen, X. S. (2020). 3D UAV trajectory design and frequency band allocation for energy-efficient and fair communication: A deep reinforcement learning approach. *IEEE Transactions on Wireless Communications*, 19(12), 7796–7809.
- [2] Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., & Monfardini, G. (2008). The graph neural network model. *IEEE transactions on neural networks*, 20(1), 61–80.
- [3] Shi, Y., Huang, Z., Feng, S., Zhong, H., Wang, W., & Sun, Y. (2020). Masked label prediction: Unified message passing model for semi-supervised classification. *arXiv preprint arXiv:2009.03509*.
- [4] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.