AFC기반 다중 스펙트럼 채널 접근을 위한 강화학습 알고리즘 비교

채준병, 박종인, *최계원
성균관대학교
chaejb88@skku.edu, pgi753@skku.edu, kaewonchoi@skku.edu

Comparision of Reinforcement Learning Algorithms for Multi-spectral Channel Access in AFC System

Jun Byung Chae, Jong In Park, *Kae Won Choi
Sungkyunkwan University

요 약

본 논문에서는 AFC(Automatic Frequency Coordination)를 사용하는 802.11ax를 모사한 환경에서 다중 스펙트럼 채널 접근을 위해 다양한 강화학습 알고리즘을 적용하고, 그 결과를 비교 분석한다. 다중 스펙트럼 채널 접근을 위한 강화학습 방법으로 DQN(Deep Q-Network), Actor-Critic와 PPO(Proximal Policy Optimization)를 사용하여 학습을 진행하였다. 본 연구를 통해 주파수를 효율적으로 사용하여 낭비되는 채널 자원을 줄여 통신속도의 향상에 도움이 될 수 있을 것이다.

Ⅰ. 서 론

통신기술은 놀라운 속도로 발전하고 통신기기들의 종류는 더 다양해지면서 수요가 증가하고, 무선 서비스품질에 대한 수요가 증가하고 있다.

이러한 상황에서 한정된 무선 채널 자원을 효율적으로 분배하는 문제가 대두되고 있다. 미국 FCC(Federal Communication Commission)는 한정된 무선 채널 자원의 효율적 이용을 위한 방안으로 선순위 미사용 중인 주파수 채널을 활용할 수 잇는 AFC System을 도입했다[1].

본 논문에서는 다수의 사용자에게 효과적으로 한정된 무선 채널 자원을 분배하기 위하여 강화학습 알고리즘을 적용하고, 그 결과를 비교 분석한다. 강화학습 모델을 학습하여 통신간 충돌을 최소화하며 채널 자원을 효율적으로 활용하는 통신 방식을 찾는 것이 본 연구의 목표이다.

Ⅱ. 본론

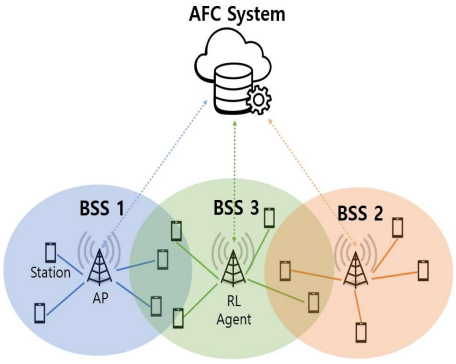
2.1 다중 스펙트럼 채널 환경

1) 환경(Environment)

강화학습 모델을 학습 시키기 위해 AFC system을 사용하는 802.11ax를 기반으로 한 다중 채널 환경을 구성하였다. 먼저 3개의 BSS(Basic Service Set)를 구성하고 각 BSS 안에서 무선 통신 환경을 구성하였다. 각 BSS에는 하나의 AP와 20개의 Station이 배치된다. BSS마다 사용할 채널을 설정하여 통신하게 되고, [그림 1]의 BSS 3 같이 하나의 BSS에 강화학습을 수행하는 Agent를 연결할 수 있다.

AFC System은 지정된 위치에서 이용 중인 선순위 사용자를 유해한 간섭으로부터 보호하는 동시에 비면허기기가 운용 가능한 주파수 대역을 자동으로 식별 후 사용 가능한 채널, 최대 송신 전력을 할당한다.

먼저 AP의 채널 선택에 있어서, 강화학습을 통해 합리적이고 효율적인 채널 선택을 할 수 있는지에 대해 확인할 수 있도록 하였다. Preamble Puncturing으로 인해 사용할 수 있는 채널이 늘어나면서, 강화학습을 통해 비어있는 채널을 효율적으로 선택해서 사용할 수 있는 알고리즘을 개발할 수 있도록 시뮬레이터를 설계하였다.



[그림 1] 무선 통신 환경 및 BSS 구조

또한, OFDMA(Orthogonal Frequency Division Multiple Access)를 통한 효율적인 채널 자원 관리에도 초점을 두고, 하나의 채널 안에서 여러 AP와 스테이션들끼리 통신하는 것도 시뮬레이터에 적용하였다. 강화학습을 통해 적절하게 RU를 지정해 줌으로써 얼마나 데이터를 효율적으로 보낼 수 있는지 확인하기 위한 부분을 적용했다.

시뮬레이터는 이벤트 발생에 따라 환경을 변화하는 Discrete Event Simulation(DES) 형태로 구현되었다. 시스템 내의 모든 AP와 station이 독립적으로 동작하고 전체 시스템이 하나의 연속된 시간에서 동작하며 Agent의 매 action마다 설정된 시간에 따라 시뮬레이션 시간이 다르게 진행된다.

2) 관찰(observation)

관찰은 채널 정보, AFC system에 대한 정보, 그리고 Data traffic 처리방식에 사용하는 Queue의 정보로 구성되어 있다.

3) 행동(action)

행동은 신호 감지 동작과 데이터 전송 동작 2가지가 있다. 신호 감지 동작은 모든 채널에서 수신된 신호를 확인하며 각 채널의 사용 여부를 확인하는 행동이다. 데이터 전송 동작은 전체 채널 중에서 사용할 채널, Station과 Packet 길이를 결정하여 데이터를 전송하는 행동이다.

4) 보상(reward)

보상은 매 step 마다 packet success, collision, packet delayed, packet dropped 4가지 지표의 가중합으로 계산 된다.

2.2 강화학습 알고리즘

본 연구에 사용되는 학습 알고리즘은 DQN, Actor-Critic과 PPO 알고리즘이다.

1) DQN(Deep Q-Network)

DQN은 Q-Network기반에 심층학습을 적용한 기술로 기존 Q-Network의 문제점을 보완하기 위해 Experience Replay와 Target Network 알고리즘이 사용되었다[2]. Q-Network는 상태 집합으로부터 행동을 결정하는 Q-function을 인공신경망으로 구성하여 신경망 내의 가중치를 학습하여 최적의 행동을 선택하는 방식이다. 상태가 신경망의 입력값으로 들어가면 상태에서 가능한 모든 행동에 대한 보상의 예측값을 얻게 되고 예측된 보상 값을 목표 보상 값 y와의 오차를 계산하는 함수인 가 최소값에 수렴하도록 가중치를 경사하강법(gradient descent)를 이용하여 최적화하였다.

2) Actor-Critic

Actor-Critic은 정책을 담당하는 Actor 네트워크와 평가하는 Critic 네트워크로 구성되어 있다. Actor 네트워크는 Policy Gradient를 통해 정책 함수를 추정하고, Critic 네트워크를 통해 시간차 학습(Temporal Difference learning) 방법으로 에피소드 내의 매 단계마다 네트워크의 파라미터를 업데이트하여 행동 값 함수를 추정한다.

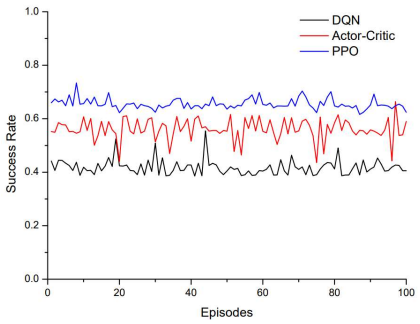
3) PPO(Proximal Policy Gradient)

PPO 알고리즘은 기본 구조로 Actor-Critic 구조를 사용하고 있다. 새로운 손실 함수(Loss Function)인 대리 함수(Surrogate Function)을 정의한다. 여기서 Policy Gradient에서 사용하는 로그 함수를 통해 표현한 목표 함수 대신에 현재의 정책과 이전의 정책의 비율을 통해 대리 함수를 정의한다. 또한 클리핑(Clipping) 방법을 사용하여 정책 업데이트 시에 정책의 변화가 크지 않게 한다[3].

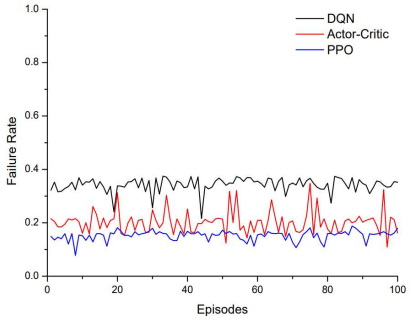
Ⅲ. 구현 및 결과

본 연구에 사용되는 학습 알고리즘 DQN, Actor-Critic과 PPO 알고리의 결과는 [그림 2], [그림 3]과 같다. 두 그래프는 매 에피소드 별 패킷 전송 성공 확률 및 실패 확률에 관한 결과이다. [그림 2], [그림 3]에서 확인할 수 있듯이 PPO 알고리즘이 DQN, Actor-Critic 알고리즘과 비교하여 높은 성공 확률, 낮은 실패 확률로 패킷을 전송함을 확인할 수 있다.

이렇게 다양한 강화학습 알고리즘을 시뮬레이터에 적용하여 각 알고리즘의 성능을 비교하였고, 이를 통해 시뮬레이터가 잘 동작함을 검증하였다.



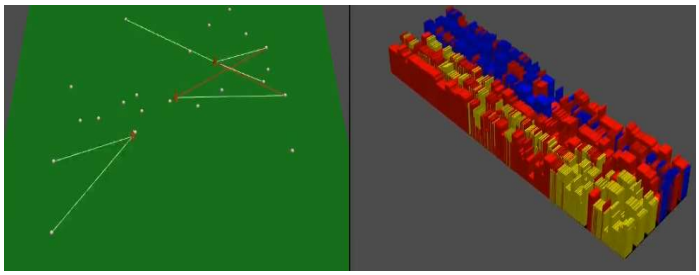
[그림 2] Success Rate 비교



[그림 3] Failure Rate 비교

[그림 4]는 학습된 강화학습 모델이 채널 접근하여 통신한 모습을 시각화한 것이다. [그림 4] 왼쪽을 보면 각각의 Station마다 다른 채널을 사용함을 색으로 구분하여 확인할 수 있도록 하였다. 여기서 서로 다른 AP가 같은 Station과 통신하면 빨간색 선으로 간섭이 일어나고 있음을 표시하였다.

또한, [그림 4]의 오른쪽을 보면, 채널별로 얼마만큼 동안 얼마만큼의 세기로 전송했는지 확인할 수 있다. AP끼리는 색깔별로 구분해서 쉽게 눈으로 구별할 수 있다. 들어온 채널의 개수만큼 필드를 나누고, 채널마다 AP가 언제 전송했는지 확인할 수 있어서 현재의 통신 진행 상황을 알 수 있도록 시각화하였다.



[그림 4] BSS 통신 모습 시각화 및 주파수 채널 사용 시각화

Ⅳ. 결론 및 향후 연구 방향

본 논문에서는 다중 스펙트럼 채널 환경에서 다양한 강화학습을 활용하여 현재 채널 상황에 맞춰 채널에 접근하여 통신하는 기술을 비교 분석하였다.

추후 연구에는 Model-based 강화학습을 사용하여 이전보다 복잡하고 다양한 환경에서도 채널 자원을 효율적으로 활용하는 통신 방식으로 확장하겠다.

ACKNOWLEDGMENT

이 논문은 4단계 BK21 사업의 지원을 받아 수행된 연구임.

참 고 문 헌

[1] FCC, Notice of proposed rulemakiing; In the matter of unlicensed use of the 6 GHz band; expanding flexible use in mid-band spectrum between 3.7 and 24 GHz, 2018.

[2] V. Mnih et al.,"Playing Atari with Deep Reinforcement Learning", Google DeepMind, 2016.

[3] Schulman. J, Wolski. F, Dhariwal. P, Radford. A, Klimov. O, “Proximal Policy Optimization Algorithms”, ArXiv Abs/1707.06347, 2017