

이웃 셀 간섭을 고려한 심층 강화학습 기반 최적 빔 선택 기법

류시아¹, 김덕경^{1*}
인하대학교¹ryusia98@daum.net¹, kdk@inha.ac.kr^{1*}

Deep Reinforcement Learning based Optimal Beam Selection by Considering Interference from a Neighbor cell

Si A Ryu¹, Duk Kyung Kim^{1*}
Inha University¹

요 약

기존 연구에서는 심층 강화학습을 적용하여 최적의 빔을 선택하는 방식을 제안하였지만, 빔 선택 시 최적의 빔에 머무르는 행동을 하지 않고, 초기 빔을 랜덤으로 설정한다는 점에서 학습 결과가 최적으로 수렴하지 않는다는 단점이 존재한다. 본 논문에서는 이를 보완하여 행동에 'stay'를 추가하고 초기 빔을 단말의 위치를 고려하여 선택함으로써 학습의 수렴성을 높인다.

I. 서론

무선 통신 시스템에서 5G의 등장으로 빠른 데이터 전송 속도에 대한 수요가 늘고 있다. 기지국에서의 적절한 빔 선택을 통한 인접 셀간 간섭 조정으로 단말의 SINR(Signal to Interference plus Noise Ratio)이 타겟을 만족하도록 하여 결과적으로 데이터 전송 속도를 향상시킬 수 있고, 재전송을 방지할 수 있다.

최근 많은 연구에서 인접 셀간 간섭 조정을 위하여 심층 강화학습 프레임워크를 기반으로 한 온라인 학습 방법을 사용하고 있다.[1] 실시간으로 얻을 수 있는 데이터를 이용하여 빠른 변화에 스스로 적응할 수 있는 온라인 학습 방법은 단말의 위치, 사용하는 빔 등에 따라 실시간으로 변화하는 무선 통신 환경에 적용하기 적합하다. SINR과 채널 합 전송률(sum rate)을 최대화하기 위하여 이러한 프레임워크를 무선 통신 시스템에 적용할 수 있다.

이와 관련하여 기존 연구에서는 빔포밍과 간섭 조정을 동시에 고려하기 위한 방법에 대하여 제시하였다.[2] 해당 연구에서는 서빙(serving) 기지국뿐만 아니라 간섭(interfering) 기지국의 영향을 함께 고려하였다. 또한 이 과정에서 심층 강화학습을 적용함으로써 모든 빔에 대하여 다 탐색해야 할 필요성을 없앴고, 여러 개의 기지국이 존재하는 환경에서의 알고리즘 구현 복잡도를 감소시켰다. 강화학습을 구현할 때는 공동행동(joint action)을 이용하여 여러 개의 기지국의 행동 선택이 한 번에 이루어질 수 있도록 하였다.

그러나 기존 연구에서는 빔을 선택하는 행동을 빔포밍 코드북 인덱스 1 증가, 1 감소 두 가지로 제한하였다. 그렇기 때문에 최적의 빔을 찾은 경우에도 해당 빔에 계속 머무르지 못하여 하나의 값에 수렴하지 못하는 한계가 존재한다. 또한 기존 연구에서는 초기 빔을 빔포밍 코드북에서 랜덤하게 선택한 값을 이용하여 결정한다. 그러나 기존 신호를 이용하여 단말이 위치한 방향의 비교적 강한 빔을 선택한 후 수신 SINR, 단말의 이동성 등을 고려하여 해당 빔을 중심으로 탐색하는 것이 일반적으로 빔 선택에서 사용하는 방법이다. 따라서 초기 빔을 랜덤으로 선택하는 것은 최적의 빔을 찾는 데에 한계가 존재한다. 또한 초기 빔으로 전혀 다른 반대 방향의 빔을 선택하는 경우가 발생하기 때문에 학습이 제대로 수렴하지 않을 수 있다는 단점이 존재한다.

본 논문이 제시하는 방법은 이러한 기존 연구의 한계

를 보완한다. 빔 선택 과정에서 행동에 'stay'를 추가함으로써 최적 빔을 선택했을 경우 해당 빔을 계속 선택할 수 있도록 한다. 또한 초기 빔을 단말의 위치를 고려하여 선택한다. 단말이 위치한 방향에 가장 가까운 방향을 가지는 빔포밍 벡터를 이용하여 생성한 빔을 초기 빔으로 선택한다.

제안 방식을 이용함으로써 강화학습의 수렴성이 커질 것이며 이를 통해 간섭이 존재하는 무선 통신 환경에서 최적의 빔을 선택하도록 학습할 수 있다. 결과적으로 단말들의 수신 SINR과 채널 합 전송률을 크게 할 수 있다.

II. 본론

1) 무선 네트워크 구조 및 환경

본 논문에서는 기지국 2 개, 단말 2 개로 구성된 네트워크 구조를 가정한다. 단말은 각각 1 개의 기지국으로부터 원하는 신호를 수신하고 다른 1 개의 기지국으로부터 간섭 신호를 수신한다. 이 때, 한 기지국이 어떤 단말에게는 서비스 기지국이지만 다른 단말에게는 간섭 기지국인 관계가 이루어진다. (그림 1) 기지국은 ULA(Uniform Linear Array) 안테나, 단말은 단일 안테나의 사용을 가정한다. mmWave 대역에서의 상황을 고려하며, 이 때 중심 주파수가 크기 때문에 전파 손실의 영향을 완화하기 위하여 아날로그 빔포밍 사용을 가정한다. 빔은 빔포밍 코드북으로부터 선택된 빔포밍 벡터를 이용하여 형성된다.

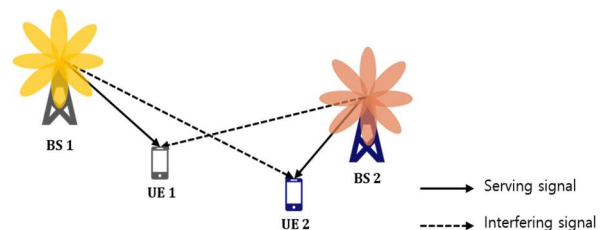


그림 1. 무선 네트워크의 구조

2) DQN (Deep Q Network)

강화학습은 에이전트가 환경으로부터 제공된 상태를 기반으로 보상을 최대화하기 위하여 행동을 선택하고 이러한 경향을 학습하는 과정을 말한다. 강화학습에 딥러닝

을 결합한 것이 심층 강화학습이며, 본 논문에서는 그 중 대표적인 구조인 DQN 을 사용한다. 매 에피소드마다 위 과정을 반복하여 학습이 진행되며 DQN 의 기본 구조는 그림 2 와 같다.

제안 방식에서 사용하는 강화학습의 요소는 다음과 같다.[3]

- 상태: 두 단말의 x 좌표, y 좌표 값과 두 기지국이 각각 선택하는 빔 인덱스를 사용한다.
- 행동: 기존 방식의 경우 코드북 인덱스 1 증가 또는 코드북 인덱스 1 감소 두 가지의 행동만을 사용한다. 반면, 제안하는 수정 방식의 경우 현재의 빔포밍 코드북 인덱스를 다시 선택하는 'stay' 행동을 추가하였다. 이를 토대로 정의한 공동행동 공간(joint action space)의 크기는 $4(2 \times 2)$ 에서 $9(3 \times 3)$ 로 증가한다.
- 보상: 두 단말에서 각각 계산된 수신 SINR 의 합으로 정의한다.
- 종료 조건: 각 단말에서의 수신 SINR 값이 정의된 최소 SINR 값 이하이면 해당 에피소드가 중지되어 종료된다.

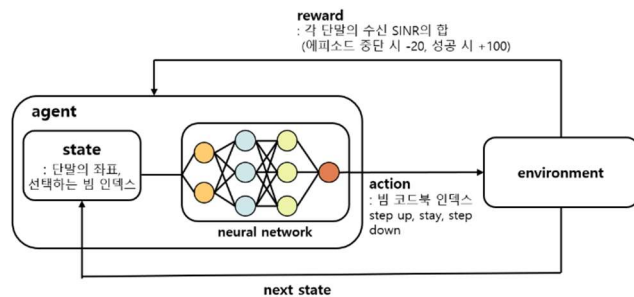


그림 2. DQN(Deep Q Network)의 기본 구조

3) 제안하는 방식

본 논문에서 이루고자 하는 바는 간섭을 고려한 무선 통신 환경에서 타겟 수신 SINR 값을 만족하는 빔을 찾는 것이다. 이에 대하여 기존 방식에서 존재하는 한계를 해결하기 위해 기존 방식을 보완하여 새로운 방식을 제시하였다.

제안 방식에서는 'stay' 행동을 추가함으로써 최적의 빔을 선택했을 경우 해당 빔 인덱스에 머무를 수 있도록 하였다. 또한, 초기 빔 인덱스를 랜덤하게 선택하는 기존 방식과 달리 제안 방식에서는 단말의 위치를 반영하여 그 위치와 가까운 빔포밍 벡터를 가지는 빔을 초기 빔으로 선택하도록 하였다. 기존 방식에서의 개선을 통하여 기존 방식과 비교하였을 때, 제안 방식에서 최적의 빔을 선택하는 방향으로 더 잘 수렴하는 것을 목표로 한다.

시뮬레이션을 진행할 때 다음과 같은 과정을 거친다. 첫 번째로 에피소드가 시작되고, 초기 상태 및 행동이 결정된다. 이 때 단말의 위치는 랜덤하게 선택되며 초기 빔 인덱스는 앞서 정의한 것처럼 단말의 위치에 따라 선택된다. 두 번째로 한 에피소드 내에서 최대 10 번의 스텝이 진행된다. 이 때 앞서 정의한 종료 조건을 만족할 경우 해당 에피소드는 중지되며, 그렇지 않고 10 번의 스텝 동안 진행이 됐을 때 최종 10 번째 스텝에서의 수신 SINR 값이 타겟을 만족할 경우 성공한 에피소드로 간주한다. 해당 에피소드가 중지되거나 성공하면 다음 에피소드로 넘어가고 이 과정을 반복하며 학습이 진행된다.

이 때 성공한 에피소드의 횟수를 관찰하여 학습의 수렴 여부를 확인할 수 있다. 본 논문에서 사용되는 알고리즘은 이러한 성공한 에피소드의 횟수를 늘리는 방향, 즉 타겟 SINR 을 만족하는 빔을 제대로 찾도록 하는 방향으

로 학습을 진행하는 것이다.

4) 시뮬레이션 결과

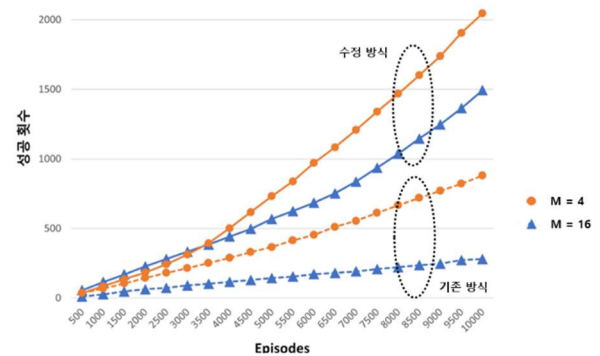


그림 3. 에피소드 진행에 따른 누적 성공 횟수

기지국의 안테나 개수 $M = 4, 16$ 에 대하여 시뮬레이션을 진행하였다. 그림 3 은 시뮬레이션을 통해 얻은 에피소드 진행에 따른 누적 성공 횟수를 나타낸다.

안테나 개수 $M = 4$ 인 경우, 제안된 수정 방식에서 성공 횟수가 2046 번으로 기존 방식에서의 성공 횟수인 881 번에 비하여 높은 성공 확률을 가지는 것을 확인할 수 있다. 또한 그래프의 기울기가 에피소드가 진행될수록 점차 커지는 것을 보아 학습 후반으로 갈수록 제대로 학습되어 수렴하고 있음을 알 수 있다.

안테나 개수 $M = 16$ 인 경우에도 $M = 4$ 인 경우와 유사한 결과가 나타난다. 다만 안테나 개수가 늘어남에 따라 선택할 수 있는 빔의 수도 늘어나기 때문에 $M=4$ 인 경우에 비하여 수렴성이 떨어진다.

총 에피소드 횟수(10000 에피소드)에 비해 성공 횟수가 비교적 적은 이유는 단말의 위치가 랜덤하게 선택될 때 어떤 빔을 선택하든 타겟 수신 SINR 을 만족하지 않는 경우가 발생하기 때문인 것으로 추측한다.

III. 결론

본 논문에서 제안하는 방식의 시뮬레이션 결과에 따르면, 기존 방식에 비해 제안 방식이 성공 확률이 높고, 학습 성능도 더 좋게 나타난다. 즉, 심층 강화학습을 이용하여 빔포밍과 간섭 조정을 고려하며 단말의 수신 SINR 과 채널 합 전송률을 증가시킬 수 있고, 결과적으로 5G 환경에서 데이터 전송 속도를 향상시키고 재전송을 줄일 수 있을 것이다.

ACKNOWLEDGMENT

본 연구는 ETRI 의 지원으로 수행되었음 (No. 2021-0-00746, Tbps 급 무선통신 기술 개발)

참 고 문 헌

- [1] M.Bennis and D.Niyato, "A Q-learning Based Approach to Interference Avoidance in Self-Organized Femtocell Networks," in Proc. IEEE Globecom Workshops, Dec.2010.
- [2] Faris B, Senior Member, IEEE, Brian L.Evans, Fellow, IEEE, and Ahmed Alkhateeb, Member, IEEE, "Deep Reinforcement Learning for 5G Networks: Joint Beamforming, Power Control, and Interference Coordination", IEEE Transactions on Communications, March, 2020.
- [3] R. S. Sutton and A.G. Barto, Introduction to Reinforcement Learning. The MIT Press, 1998.