

MIMO 시스템을 위한 Actor-Critic 심층강화학습 기반 안테나 선택 기법

오영우, 최우열*

조선대학교

snlyoungwoo@gmail.com, *wyc@chosun.ac.kr

An Actor-Critic Deep Reinforcement Learning-Based Antenna Selection Scheme for MIMO Systems

Youngwoo Oh, Wooyeol Choi*

Chosun University

요약

MIMO 시스템에서의 안테나 선택 기법은 하드웨어 비용을 줄임과 동시에 최대 채널 용량을 얻기 위한 기술이다. 기존 안테나 선택 기법 중 최적의 성능을 제공하는 brute-force search 기반의 안테나 선택은 가능한 모든 안테나 부집합에 대한 채널 용량을 비교함으로써 높은 계산 복잡도를 요구한다. 따라서, 본 논문에서는 최적 방식에서 제기되는 문제를 해결하기 위해 심층강화학습을 활용한다. 제안하는 방식은 actor-critic (A2C) 알고리즘을 활용하여 행동 확률에 대한 상태 가치를 평가함으로써 효율적으로 최적의 안테나 부집합을 예측한다. 실험 결과를 통해, 제안하는 심층강화학습 기반 안테나 선택 기법이 brute-force search 기법과 근접한 성능을 제공할 수 있음을 확인하였다.

I. 서론

Multiple-input multiple-output (MIMO) 시스템은 다중 안테나를 활용함으로써, 높은 채널 용량을 달성할 수 있다. 그러나 동시에 RF chain의 수가 증가함에 따라 하드웨어 복잡도 및 시스템 비용이 증가하는 문제를 초래한다. 이러한 문제를 해결하기 위해 최근 다양한 안테나 선택 기법의 연구가 활발히 진행되고 있다 [1-3]. 그러나, 대부분 안테나 선택 기법이 행렬 연산과 조합식을 통해 최적의 부집합을 선택하는 brute-force search, greedy 알고리즘을 기반으로 한다 [4]. 이러한 방식은 최적의 성능을 제공하지만, 안테나 개수에 따라 복잡도와 성능 간 trade-off 되는 문제를 지녔다. 따라서 본 논문에서는 이러한 문제를 해결하기 위해 deep reinforcement learning (DRL)을 활용한 안테나 선택 기법을 제안한다.

II. 시스템 및 채널 모델

본 논문에서는 M 개의 송신 안테나와 N 개의 수신 안테나로 구성된 MIMO 시스템을 고려하였다. 이때, 채널은 Rayleigh fading으로 가정하였으며, 해당 시스템의 신호 모델은 다음과 같이 나타낼 수 있다.

$$r = \sqrt{\frac{S}{M}} Hx + n \quad (1)$$

수식 1의 r, x, n 은 각각 수신된 신호, 전송 신호, 그리고 가우시안 잡음을 의미한다. H 는 송·수신 안테나로 구성된 $M \times N$ 의 채널 행렬, S 는 신호 대 잡음 비율의 평균을 나타낸다. 이때, j 번째 안테나 부집합 채널 $\bar{H}_{(J)}$ 에 대한 채널 용량은 수식 2와 같이 표현된다.

$$C(\bar{H}_{(J)}) = \log_2 \det \left(I_N + \frac{S}{M} \bar{H}_{(J)} \bar{H}_{(J)}^H \right) \quad (2)$$

M 개의 안테나 중 L 개의 안테나 선택을 가정할 때, 가능한 모든 부집합의 수는 $N_c = \binom{M}{L}$ 로 구성된다. 이때, brute-force search 기반 안테나 선택은 N_c 개의 안테나 조합 중 수식 3에 의해 채널 용량을 최대로 하는 안테나 부집합 인덱스 J^* 를 선택하게 된다.

$$J^* = \underset{J \in N_c}{\operatorname{argmax}} C(\bar{H}_{(J)}) \quad (3)$$

이러한 방식은 모든 안테나 조합을 탐색하여 최대 채널 용량을 획득할 수 있으나, 안테나 개수에 따라 계산 복잡도가 기하급수적으로 증가한다.

III. 심층강화학습 기반 안테나 선택 기법

1. Markov Decision Process(MDP) 정의

본 논문에서는 brute-force search 기반 안테나 선택 기법에서 제기되는 문제점을 해결하기 위해 심층강화학습을 활용한다. 이를 적용하기에 앞서, 다음과 같은 상태, 행동, 보상에 대해 정의한다. 행동 a 는 수식 4와 같이 안테나 조합 N_c 로 구성된다. 상태 s 는 수식 5와 같이 행동에 대한 채널 용량 $C([a_t])$ 로 구성된 유한 집합으로 정의된다. 이때, 행동이 변화하는 매 스텝 t 에 따라 산출되는 상태는 가능한 모든 상태의 집합 S 에 포함된다.

$$a = (N_1, \dots, N_c), \quad a \in A \quad (4)$$

$$s = (C([a_t]), C([a_{t+1}]), \dots), \quad s \in S \quad (5)$$

상태와 행동에 대한 보상 메커니즘은 수식 6과 같이 구성된다. 이때, s_{\max} 는 상태 집합 S 의 최댓값을 의미한다.

$$r[t] = \begin{cases} 1, & \text{if } s_t \geq s_{\max} \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

상태 및 행동에 대해 예상되는 장기 보상은 다음과 수식 7과 같이 표현할 수 있으며, 안테나 부집합 선택 문제는 수식 8, 9와 같이 심층강화학습으로 해결 가능한 MDP 문제로 변환된다.

$$R = E[r[t] + \alpha r[t+1] + \alpha^2 r[t+2] + \dots | S = s, A = a] \quad (7)$$

$$\underset{a}{\operatorname{maximize}} \sum_{n=1}^N R \quad (8)$$

$$\text{s.t. } a \in A \quad (9)$$

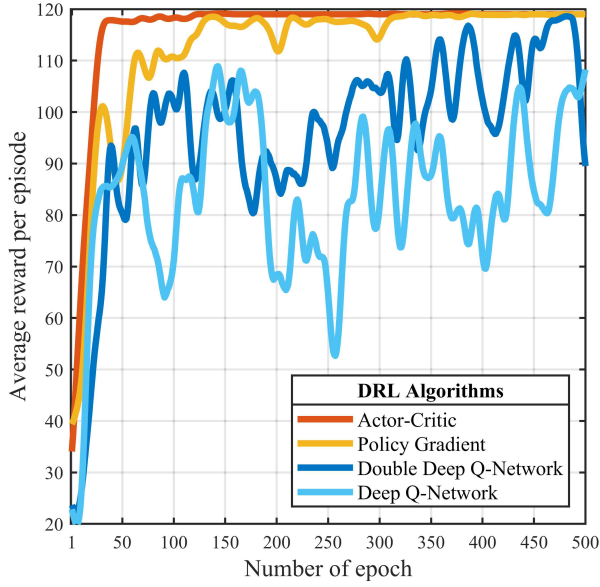


그림 1. 심층강화학습 알고리즘별 학습 성능 분석

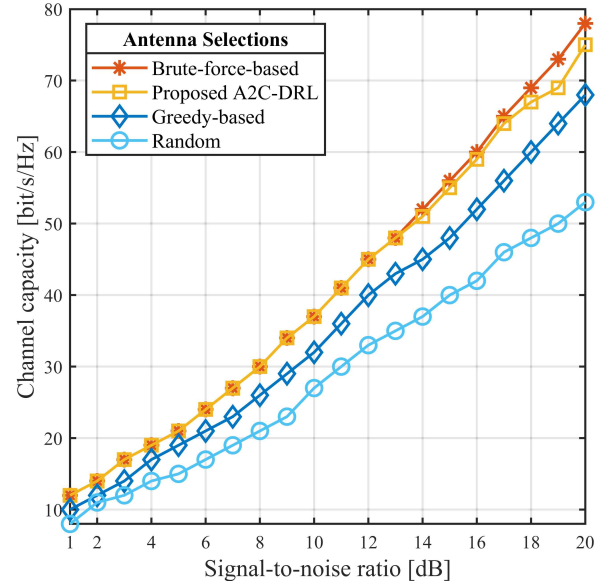


그림 2. 안테나 선택 기법에 따른 채널 용량 분석

2. Actor-critic (A2C) 설계 및 활용

본 논문에서 활용되는 A2C 알고리즘은 정책 신경망과 가치 신경망으로 구성되며, 각각의 에이전트는 두 개의 deep neural network (DNN)으로 구성된다. 사용된 DNN 모델의 경우, 각각 32, 64의 크기를 지닌 2개의 은닉계층, 그리고 출력계층으로 구성되며 각 은닉층은 rectified linear unit(ReLU) 함수에 의해 활성화된다.

Actor 에이전트의 optimizer는 Adam 함수를 사용하고, 손실함수로 categorical cross-entropy를 사용한다. 이때, 마지막 출력계층에서는 softmax 활성화 함수가 사용되어 입력되는 상태 s 에 대한 행동 확률을 산출한다. Critic 에이전트의 경우, mean squared error (MSE) 손실함수를 사용하며, actor 에이전트에서 산출된 행동 및 상태 가치를 평가한다. 이러한 과정을 통해, brute-force search 기반 안테나 선택 기법의 단점인 높은 시간 복잡도를 해결함과 동시에, 최적의 채널 용량을 지닌 안테나 조합 인덱스 J^* 를 예측할 수 있다.

IV. 성능 비교 및 분석

본 연구는 A2C 알고리즘 기반 심층강화학습을 활용한 안테나 선택 기법을 제안하였다. 제안하는 안테나 선택 기법의 성능 비교에 앞서 A2C 알고리즘을 포함한 다양한 심층강화학습 알고리즘별 학습 성능에 대한 학습 성능에 대한 분석을 수행한다.

그림 1은 심층강화학습 알고리즘에 따른 학습 성능과 수렴률에 대한 성능 분석 결과를 보여준다. DQN, DDQN 방식은 epoch가 증가함에 따라 평균 보상이 증가하지만, fluctuation 문제가 발생하는 것을 확인할 수 있다. 그에 반해, A2C, policy gradient 방식은 Q-value 기반 학습 방식에 비해 빠른 수렴속도를 보임과 동시에 안정적인 학습 성능을 보인다.

그림 2는 신호 대 잡음 비가 증가할 때, 송신 안테나 선택 기법에 따른 채널 용량의 성능을 분석한 결과를 나타낸다. 이때, 안테나 개수는 $N_t = N_r = 16$, $L = 14$ 와 같이 설정하였다. 제안하는 A2C-DRL 기법은 brute-force 기반 안테나 선택 기법과 거의 근접한 성능을 획득하였음을 보인다. 그 외, 채널 이득 값을 기반으로 안테나를 선택하는 greedy 기반 안테나 선택과 무작위 선택 순으로 높은 성능을 보인다.

V. 결론

본 논문에서는 기존 brute-force search 기반 안테나 선택 기법에서 제기되는 문제를 해결하기 위해 A2C 알고리즘 기반 심층강화학습을 활용하였다. 실험 결과를 통해, 제안한 선택 기법은 타 기법들에 비해 월등히 높은 채널 용량을 획득할 수 있으며, 높은 시간 복잡도를 가진 brute-force search 기반 안테나 선택 기법과 근접한 성능을 보임을 확인하였다. 추후 후속 연구에서는 해당 연구 결과를 기반으로 거대 다중 안테나 시스템에서 다중 사용자를 고려한 안테나 선택 기법 연구를 수행하고자 한다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학지원사업의 연구결과로 수행되었음(2017-0-00137).

참고 문헌

- [1] 서동민, 손혁민, "MIMO 시스템을 위한 저 복잡도 송신 안테나 선택 알고리즘", *한국통신학회 논문지(JKICS)*, Vol. 44, No. 6, pp. 278-280, 2019년 2월.
- [2] M. O. K. Mendonça, P. S. R. Diniz, T. N. Ferreira and L. Lovisolo, "Antenna Selection in Massive MIMO Based on Greedy Algorithms" *IEEE Transactions on Wireless Communications*, Vol. 19, No. 3, pp. 1868-1881, Mar. 2020.
- [3] R. Chataut and R. Akl, "Massive MIMO systems for 5G and beyond Networks-Overview, Recent Trends, Challenges, and Future Research Direction", *Sensors*, Vol. 20, No. 10, May. 2020.
- [4] Y. Gao, H. Vinck and T. Kaiser, "Massive MIMO Antenna Selection : Switching Architectures, Capacity Bounds, and Optimal Antenna Selection Algorithms", *IEEE Transactions and Signal Processing*, Vol. 66, No. 5, pp. 1346-1360, Mar. 2018.