Deep Reinforcement Learning 을 이용한 Adaptive Modulation selection 기법

김재익, 황상원, 이인규
고려대학교

jaeik@korea.ac.kr, tkddnjs3510@korea.ac.kr, inkyu@korea.ac.kr

A method of Adaptive Modulation selection using Deep Reinforcement Learning Approach

Jaeik Kim, Sangwon Hwang, Inkyu Lee
Korea Univ.

요 약

본 논문은 Single Cell 무선통신 환경에서 채널 변화에 따른 Adaptive Modulation selection 기법에 대해 기술한다. 채널 응답의 변화에 따라 유효한 변조 방식을 선택하는 문제에 대해 Deep Reinforcement Learning 을 통해 능동적으로 대응하여 기존 기법 대비 Spectral Efficiency 관점에서 이득이 있음을 보인다.

Ⅰ. 서 론

다양한 유무선 통신 시스템에서 채널 상태에 따라 변조 기법을 달리하여 적용하는 방식들이 널리 사용되고 있다. 무선 통신에서는 더 높은 전송률을 얻기 위해 Modulation order 와 code rate 을 조정하는 Adaptive Modulation and Coding(AMC) 기법이 널리 사용되고 있다. 일반적으로 CQI 를 활용하여 채널 응답 구간별로 적합한 Modulation and Coding Scheme(MCS)를 look-up table 에 의존한 방식으로 선택하게 되는데, 급격한 채널 응답의 변화나 시스템의 비 선형적 요소에 의한 변화에 유연하게 대처하기 어려운 단점이 있다.

현재 Deep Learning (DL)을 활용한 성능 향상 기법이 무선통신을 비롯한 여러 분야에서 다양하게 연구되고 있다. 그 중에서도 Reinforcement Learning(RL)은 off-line training 과정과 학습을 위한 data-set 이 필요하지 않아 MCS 선택 문제에 있어 상술한 단점을 해결하는데 적합하여 RL 을 적용하고자 하는 연구들이 진행되어 왔다. 그러나 Colored Noise 와 같은 제한적인 상황에서만 성능 이득을 보이거나[1], Monte Carlo run 을 통한 성능 제시에 그쳤다.[2]

본 논문에서 적용한 Deep Q-Network(DQN)을 통한 방식은 일반적인 Q-table 방식보다 우수한 학습 성능을 보이는 것으로 알려져 있으며, RL 의 장점에 더불어 경험 리플레이를 통해 최근 샘플에 과적합되지 않고 학습이 가능한 방식이다. 이를 Modulation Order selection 에 적용하고 Spectral Efficiency 관점에서 더 나은 선택을 학습시키고 기존의 table-base 방식 대비 우수한 성능을 가짐을 확인하였다.

Ⅱ. 본 론

2.1 시스템 모델

본 논문에서는 M 개의 Tx Antenna 를 갖는 MISO Single Cell 환경에서 28GHz 대역의 center frequency 를 갖는 multi path 시스템을 고려한다.

$$y_t = H_t s_t + z_t$$

여기서 z_t 는 AWGN noise 신호이며, 수신 채널은 다음과 같이 표현된다.

$$H_t = \sqrt{\rho}\sum_{i=0}^{S-1} v_{UE}(\phi_{i,t}^{UE}\theta_{i,t}^{UE}) v_{BS}(\phi_{i,t}^{BS}\theta_{i,t}^{BS}) e^{j2\pi f_i t T_s}$$

ρ 는 채널 모델의 path loss 로 ETSI 의 UMa NLOS 모델을 사용했고, S 는 전체 multi path 의 수, 그리고 $\phi_{i,t}, \theta_{i,t}$ 는 각각 UE 와 BS 의 path 별 방위각과 고도각이다. 또한, f_i 는 i 번째 path 의 doppler frequency, T_s 는 symbol period 를 나타내며, 방위각과 고도각에 따른 multi antenna 의 ULA (Uniform Linear Array) 벡터는 다음과 같이 표현된다.

$$v_{BS}(\phi_{i,t}^{BS}\theta_{i,t}^{BS}) = \frac{1}{\sqrt{M}}\begin{bmatrix} 1, e^{(j\frac{2\pi d}{\lambda}(\sin\phi_{i,t}^{BS}\sin\theta_{i,t}^{BS}+\cos\theta_{i,t}^{BS}))}, \\ \ldots, e^{(j(M-1)\frac{2\pi d}{\lambda}(\sin\phi_{i,t}^{BS}\sin\theta_{i,t}^{BS}+\cos\theta_{i,t}^{BS}))} \end{bmatrix}$$

상술한 채널 모델에 따라, fading 과 Doppler effect 가 존재하는 MISO 환경에서 UE mobility 의 영향까지 반영된 채널 응답에 적합한 Modulation Order 를 선택하는 문제가 된다.

2.2 Deep Q-Network

적절한 Modulation 을 적응적으로 학습하기 위해 DQN 을 적용하였다. DQN 의 Markov Decision Process(MDP)는 다음 수식과 같이 적용된다.[4]

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha(R_{t+1} + \gamma \max_{a'} Q(S_{t+1}, a') - Q(S_t, A_t))$$

정의된 MDP 에 대해 State(S_t)는 매 전송에 대한 CQI 응답을 [-5,40](dB)구간에 대해 15 등분하여 적용하고, 선택 가능한 Action(A_t)는 QPSK, 16QAM, 64QAM, 256QAM 의 4 가지로 구성하였다. α 는 learning rate, γ 는 discount factor 로써 표 1 에 simulation setting 이 기술되어 있다.

DQN reward 로 Spectral Efficiency(SE)를 다음과 같이 정의한다.

$$SE = (1 - E_s)Q_m$$

E_s 는 심볼 에러, Q_m 은 modulation order 별 bit 수로, 높은 modulation order 로 전송이 성공할 때 더 높은 reward 를 제공하여 DQN 이 되도록 에러 없이 많은 bit 를 전송하도록 학습하게 돕는다.

Parameter	Value
Transmit antenna	64
Receiver antenna	1
Frequency	28GHz
Transmit power	25dbm
Noise power	-123.85dbm
Cell Radius	25~90m
Azimuth angle	$[-60^{\circ}, 60^{\circ}]$
Elevation angle	$[60^{\circ}, 120^{\circ}]$
Path loss	UMa NLOS
Shadowing standard deviation	6dB
Number of path	10
Learning rate	$1e^{-3}$
Discount factor(γ)	0.99
Maximum exploration rate($\varepsilon_{\max}$)	1.0
Minimum exploration rate($\varepsilon_{\min}$)	0.1
Exploration decay rate(ε_{decay})	0.9999

표 1. Simulation parameters

2.3 모의실험

상술한 채널 모델에 대하여 UE 는 속도 v 로 Cell Edge 방향으로 등속 직진 이동한다고 가정하였으며, 상세한 시스템 파라미터들은 표 1 에 요약되어 있다.

제안한 방식의 성능 평가를 위한 비교 scheme 으로써 table-base 방식의 modulation selection 을 적용했다. 이 table 의 threshold 값은 SNR 별 error rate 을 고려하여 결정되었으며, 일반적인 방식으로 알려진 error rate 10% 기준으로 선택되었다.[3]

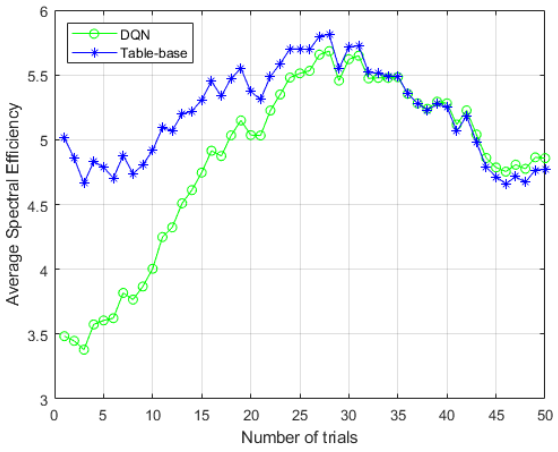

그림 1. 학습 정도에 따른 SE 추이

그림 1 은 학습이 진행됨에 따른 평균 SE 의 추이를 나타낸다. DQN 에 적용된 reward 를 기준으로 table 방식 기법을 동일하게 평가하여 성능을 비교한 결과이다. 학습이 시작된 시점에서 table-base 방식이 더 우수한 성능을 갖지만, 점차 DQN 을 통한 학습이 진행될수록 평균 SE 가 개선되어 table-base 방식 대비 우수한 성능을 보임을 확인할 수 있다.

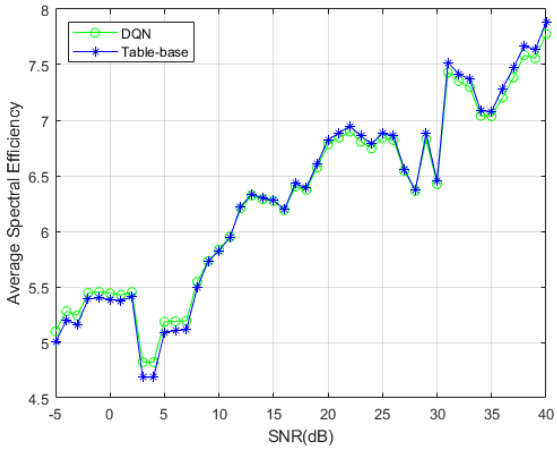

그림 2. 학습 이후 SNR(dB)에 따른 평균 SE

그림 2 는 학습이 진행된 이후의 SNR(dB)에 따른 평균 SE 의 분포를 나타낸다. 학습된 DQN 이 평균적으로 높은 SE 를 보이는 상황에서, 낮은 SNR 구간에서 상대적으로 높은 SNR 구간보다 나은 성능을 보임을 확인할 수 있다.

Ⅲ. 결 론

본 논문에서는 DQN 을 적용한 적응형 변조 방식 선택 기법을 소개하였으며, 학습 추이를 보이고 기존의 table-base 방식 대비 성능 개선을 검증하였다.

ACKNOWLEDGMENT

본 연구는 한국연구재단의 지원을 받아 수행되었음.

(No. 2017R1A2B3012316.)

참 고 문 헌

[1] J. P. Leite, P. H. P. de Carvalho and R. D. Vieira, "A flexible framework based on reinforcement learning for adaptive modulation and coding in OFDM wireless systems," 2012 IEEE Wireless Communications and Networking Conference (WCNC), 2012, pp. 809-814.

[2] M. P. Mota, D. C. Araujo, F. H. Costa Neto, A. L. F. de Almeida and F. R. Cavalcanti, "Adaptive Modulation and Coding Based on Reinforcement Learning for 5G Networks," 2019 IEEE Globecom Workshops (GC Wkshps), 2019, pp. 1-6.

[3] R. Fantacci, D. Marabissi, D. Tarchi and I. Habib, "Adaptive modulation and coding techniques for OFDMA systems," in IEEE Transactions on Wireless Communications, vol. 8, no. 9, pp. 4876-4883, September 2009.

[4] Mnih, V., Kavukcuoglu, K., Silver, D. et al. Human-level control through deep reinforcement learning. Nature 518, 529- 533 (2015).