5G 네트워크를 위한 심층 강화 학습 기반 자원 할당에 대한 동향 조사

오영우, 최우열*

조선대학교

oyw@chosun.kr, *wyc@chosun.ac.kr

A Comprehensive Survey of Current State-of-the-Art Trends in Resource Allocation Based on Deep Reinforcement Learning for 5G Networks

Youngwoo Oh, Wooyeol Choi*

Chosun University

요 약

심층 강화학습은 기계 학습의 한 영역으로 특정 환경 안에서 시행착오를 통해 수집되는 다양한 데이터의 패턴을 학습을 통해 최적의 정책을 찾는 방법으로 마르코프 결정 과정에 의해 수학적으로 모델링되며, 선택 가능한 행동 중 보상을 최대화하도록 하는 행동을 학습하는 방안이다. 최근에는 싱글 에이전트만으로 해결하기 어려운 문제들을 여러 개의 로컬 에이전트가 경쟁 및 협업하여 문제의 정답을 도출하는 멀티 에이전트 강화학습과 특정 시나리오에서 요구되는 다중 목표의 최적화 및 trade-off 개선을 위한 다중 목표 강화학습 등의 연구가 활발히 수행되고 있다. 이러한 접근 방안들은 5G 네트워크에 범용적으로 적용될 수 있는 솔루션으로 다양한 자원할당 분야에 활용되고 있으며, 기존의 대규모 안테나 사용에 따른 계산 복잡도 문제 및 최적화 문제 해결을 위한 방안으로 사용된다. 따라서, 본 논문에서는 이러한 5G 네트워크에서 심층 강화학습 전략을 활용한 자원할당 기술에 관한 연구를 조사하여, 해당 기술의 이점과 기술적 한계에 대해 논의하고자 한다.

Ⅰ. 서 론

5G 네트워크 및 massive Multiple-Input Multiple-Output (MIMO) 등의 시스템에서의 채널 용량, 에너지 효율성, 사용자 공정성 등을 개선하기 위해 다양한 자원할당 연구들이 수행되었다. 그러나, 이러한 전통적인 최적화 알고리즘 기반의 자원할당 기법은 높은 계산복잡도가 요구된다는 한계를 지닌다. 이에 따라, 이를 효과적으로 처리하기 위한 deep learning (DL) 기반의 다양한 자원할당 연구들이 수행되었다. DL 기술을 활용한 자원할당 기법들의 경우, 다양한 신경망을 활용함으로써, 최적의 성능을 달성함과 동시에, 계산복잡도를 크게 줄일 수 있음을 보였다. 그러나, 이는 시스템 모델이나 주요 매개변수가 변경될 때마다 새로운 학습 데이터셋이 요구됨으로 동적으로 변화하는 5G 환경에 적용하기 어렵다는 한계를 지닌다. 이러한 제한점을 완화할 수 있는 대안으로 deep reinforcement learning (DRL)을 활용한 접근 방식이 최근 크게 주목받고 있다. DRL 모델의 경우, 별도의 데이터셋과 전처리 과정 없이 환경과 에이전트 간의 상호작용을 통해, 전력 제어, 안테나 선택, 빔포밍 선택 등의 다양한 자원할당 연구에서 최적의 솔루션을 달성할 수 있다 [1]-[5].

따라서, 본 논문에서는 5G 및 massive MIMO 시스템을 고려하는 시나리오에서 다양한 심층 강화학습 전략을 활용한 자원할당 연구를 분석하고, 이에 대한 이점과 제한점을 제시함으로써, 무선통신에서의 강화학습 기반의 접근에 대한 통찰력을 제공하고자 한다.

Ⅱ. 본론

1. Deep reinforcement learning (DRL)

강화학습이란 환경, 행동, 상태, 보상과 같은 4가지 요소로 구성되는 학습 기법으로, 주로 싱글 에이전트를 대상으로 많은 연구가 수행되었다. 학습 과정에서의 에이전트는 action을 취한 후, 환경의 변화를 관측하고, 보상받는 과정을 반복 수행하면서, 정의된 보상을 최대화하는 방향으로 학습하게 된다. 이러한 DRL 기반의 자원할당 연구의 일환으로 [1]에서는 q-learning 기반의 전력 할당 기법 연구가 수행되었다. 해당 연구에서의 행동 공간은 최소 및 최대 전력 사이의 송신 전력으로 구성된 집합으로 정의된다. 이후, 에이전트가 학습 과정에서 선택한 action에 의해 얻어지는 signal-to-interference-plus-noise (SINR) 및 channel state information (CSI)를 상태 공간으로 사용하며, sum-rate를 보상으로 설정하여, 이를 최대화하는 정책이 활용된다. 반면, [2]에서는 q-table 기반으로 정책을 생성 및 갱신하는 q-learning의 학습 속도 저하 문제를 개선하기 위해, 정책 경사 기반의 강화학습을 활용하였다. 활용된 advantage actor-critic (A2C) 기반의 자원할당 기법은 action에 따라 관찰되는 state를 통해 최적의 action을 선택할 수 있도록 하는 policy network와 선택된 action을 평가하는 value network로 구성된다. 이러한 학습 정책은 actor-critic network 간의 선택 및 비평 과정을 통해, 기존의 q-learning 방식에 발생하는 느린 수렴 속도와 불안정성을 대폭 개선하였다. 그러나, 단일 에이전트를 통해 얻어지는 환경의 한정적인 데이터를 기반으로 학습이 수행되게 되므로, 환경의 모든 요소를 탐색할 수 없어 발생하는 sub-optimal 수렴과 space 크기에 따라 학습 속도가 크게 저하되는 차원-저주 문제를 지님으로, massive MIMO 및 5G 환경에 대해 최적의 성능을 온전히 보장하기 어렵다는 문제를 지닌다.

2. Multi-agent reinforcement learning (MARL)

최근 데이터를 분산하여, 저장하는 딥러닝 분산 구조를 적용하는 것에서부터 각각의 독립적인 환경에서 얻은 경험과 데이터를 빠르게 수집하기 위한 것까지 multi-agent 구조가 적극적으로 활용하고 있다. 이때, 멀티 에이전트 강화학습은 다수의 로컬 에이전트가 동시에 환경을 공유하는 상태에서 각자 또는 공통의 목표를 위해 자신의 행동 정책을 학습하는 기법을 의미한다. 멀티 에이전트 강화학습에서 다수의 에이전트는 동일한 시간과 공간을 공유하는 조건 아래에서 주어진 목표의 형태에 따라 경쟁적

또는 협업적 행동을 취하기 위해 정책을 학습함으로써, 기존의 단일 에이전트와 환경의 상호작용에서 얻어지는 데이터를 효과적으로 수집 및 학습하는 방안으로 빠른 수렴을 달성할 수 있다. [3]에서는 이러한 이점을 활용하여, 분산 학습 및 실행의 확장성을 고려한 MARL 기반의 자원할당 기술을 제안하였다. Multi-cellular 환경에서 모든 local deep q-network (DQN) agent는 동일한 target 매개변수를 지니며, 각각의 agent로부터 얻어지는 궤적(trajectory)을 활용하여, global network를 업데이트시킨다. 이러한 접근 방식은 다중 에이전트를 활용하여, 기존의 DRL 접근 방식보다 빠르고, 효과적인 학습 성능을 달성할 수 있음을 보인다. 그러나, 이러한 multi-agent의 활용은 단일 목표를 최적화하기 위한 자원할당 기술에서의 학습 속도 및 수렴 성능을 증진하는 방안으로 주로 사용되므로, 5G network 및 massive MIMO 환경에서 요구되는 주파수 효율성, 에너지 효율성, 사용자 공정성 등의 핵심 지표 간의 trade-off 문제를 해결하기 어렵다는 한계를 지닌다.

3. Multi-objective reinforcement learning (MORL)

기존의 DRL, MARL 알고리즘 발전에도 불구하고, 한 가지 남은 과제는 더 크고, 복잡한 문제를 해결하기 위해 확장하는 것으로, 이는 실세계에 발생하는 다중 목표의 최적화 문제를 해결하기 위한 MORL 알고리즘의 등장 배경이 되었다. 이때, MORL 모델은 다중 목표의 개념을 포함한 multi-objective MDP (MOMDP)에 의해 모델링된다. 나아가, DRL과 MORL 알고리즘의 가장 큰 차이는 DRL의 경우, 단일 목표의 최적화를 위한 보상으로 스칼라값이 사용되는 반면, MORL 알고리즘은 각 목표에 대한 고유 보상에 존재하므로 보상 벡터로 정의된다. 또한, 다중 목표 간의 우선순위를 결정하기 위해 그림 2와 같이 선호도 가중치가 활용된다. 최근에는 이러한 MORL 알고리즘을 활용하여, massive MIMO 환경에서 요구되는 채널 용량과 사용자 공정성과 같은 다중 목표를 동시에 최적화하기 위한 MORL 기반의 multi-objective optimization (MOO) 최적화 연구가 수행되었다 [4], [5].

[4]의 경우, A2C 알고리즘과 replay buffer를 결합하여 Pareto optimal-approximation 정책을 통해 채널 용량과 사용자 공정성을 동시에 최적화하였다. 이때, 해당 방식은 취한 행동에 따라 얻어지는 보상 벡터를 활용하여, 생성된 Pareto dominance와 frontier를 기반으로 계산되는 Pareto 최적을 선호도 가중치를 활용하여, 최적의 frontier를 근사하는 방향으로 학습하게 된다. 이때, dominance는 action을 취했을 때, Pareto optimal이 아닌 모든 값을 의미하고, frontier는 모든 dominance를 지배하는 optimal 값을 의미한다. 이와 달리, [5]에서는 기존의 DRL에서 보상을 최대화하기 위해 사용되는 linear sum 방식을 활용하여, MOO 문제를 single-objective optimization (SOO)로 축소한 뒤, twin delayed deep deterministic policy gradient (TD3) 알고리즘을 이용하여, 기존 자원할당 기법 대비 우수한 성능을 달성할 수 있음을 보였다. 그러나, 대부분의 MORL 알고리즘이 Pareto optimal approximation 방식을 따르며, 이는 massive MIMO 환경과 같이 주요 시스템 매개변수의 크기에 따라 hyper-volume 증가와 더불어, 값을 지속해서 비교하여 Pareto optimal을 결정해야 하므로 많은 학습 시간이 소요된다. 또한, linear sum 방식의 경우, 보상 벡터를 하나의 단일 스칼라값으로 처리함에 따라, 빠른 수렴 속도를 제공할 수 있으나, trade-off 관계가 성립하는 MOO 문제 해결을 보장할 수 없다. 따라서, MORL 학습 전략을 활용한 자원할당 기술은 효과적인 선호도 가중치 결정 방안과 더불어 데이터를 효과적으로 수집 및 탐험하기 위한 새로운 접근의 학습 전략이 요구된다.

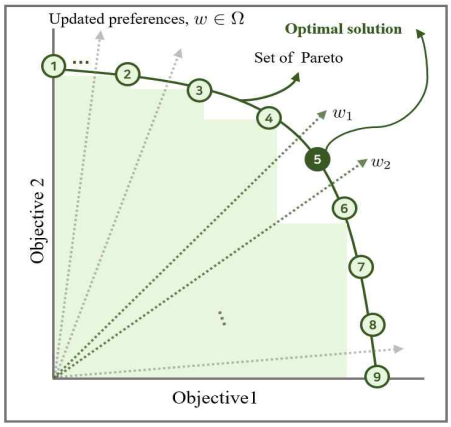

그림 1. 다중 목표 강화학습에서의 선호도 가중치 활용

Ⅲ. 결론

본 논문에서는 5G network 및 massive MIMO 시스템을 위한 심층 강화학습을 활용한 자원할당 연구를 분석하였다. 분석 결과를 통해, DRL 기술의 활용은 기존의 최적화 알고리즘 및 DL 접근 방안에서 요구되는 높은 계산복잡도 및 데이터셋 등의 문제를 해결할 수 있는 대안으로, 다양한 연구에서 DRL 기반 자원할당 기술의 우수성이 검증되었다. 나아가, 가장 최근에는 차세대 무선 통신 시스템에서 요구되는 다양한 목표를 동시에 최적화하기 위한 다중 목표 강화학습 등의 연구가 활발히 수행되고 있으며, massive MIMO에서 활용되는 대규모 안테나 및 동적으로 변화하는 채널 특성을 고려한다면, 심층 강화학습을 활용한 자원할당 연구는 더욱 활성화될 것으로 전망된다.

ACKNOWLEDGMENT

Put sponsor acknowledgments.

참 고 문 헌

[1] F. Meng, P. Chen and L. Wu,“Power Allocation in Multi-User Cellular Networks with Deep Q Learning Approach,” *in Proc. ICC*, pp. 1-6, Shanghai, China, May. 2019.

[2] L. Chen, F. Sun, K. Li, R. Chen, Y. Yang and J. Wang, “Deep Reinforcement Learning for Resource Allocation in Massive MIMO,” pp. 1611-1615, *in Proc EUSIPCO*, Dublin, Ireland, Aug. 2021.

[3] Y. S. Nasir and D. Guo, “Multi-Agent Deep Reinforcement Learning for Dynamic Power Allocation in Wireless Networks,” in IEEE Journal on Selected Areas in Communications, vol. 37, no. 10, pp. 2239-2250, Oct. 2019.

[4] R. Chen et al., “Adaptive Multi-objective Reinforcement Learning for Pareto Frontier Approximation: A Case Study of Resource Allocation Network in Massive MIMO,” pp. 1631-1635, *in Proc EUSIPCO*, Dublin, Ireland, Aug. 2021.

[5] M. Rahmani, M. Bashar, M. J. Dehghani, P. Xiao, R. Tafazolli and M. Debbah, “Deep Reinforcement Learning-based Power Allocation in Uplink Cell-Free Massive MIMO,” pp. 459-464, *in Proc WCNC*, Austin, TX, USA, Apr. 2022.