자기조직 다중에이전트 심층강화학습 기술 동향

이현수°, 백한결, 박수현, 김중헌
고려대학교

{hyunsoo, 67back, soohyun828, joongheon}@korea.ac.kr

Trends in Multi-Agent Self-Organizing Deep Reinforcement Learning

Hyunsoo Lee°, Hankyul Baek, Soohyun Park, Joongheon Kim
Korea University

요 약

본 연구는 최근 활발한 연구가 진행 중인 자기조직 강화학습과 인공지능 기술의 최신 동향에 초점을 맞추고 있다. 특히, 본 논문은 에이전트들이 원활하게 협업할 수 있도록 자율적으로 그룹을 형성하거나 정보를 교환하는 방법론에 대해 상세히 논의한다. 더불어, 다중 에이전트 강화학습에서 성능 평가의 표준으로 자리 잡고 있는 StarCraft Multi-Agent Challenge (SMAC)과 Predator Prey 시나리오에 대해서도 깊이 있게 탐구한다. 이러한 분석은 다중 에이전트 시스템의 효율적인 설계와 구현을 위한 중요한 기초를 제공한다.

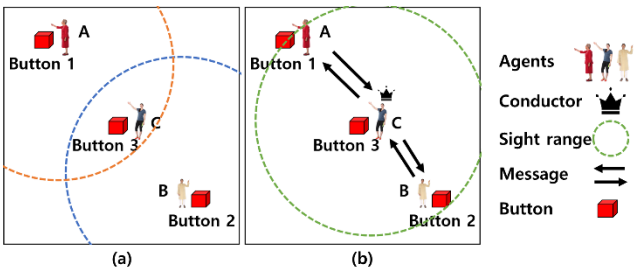
Ⅰ. 서 론

다중 에이전트 강화학습은 둘 이상의 에이전트가 서로 상호작용하면서 문제를 해결하는 강화학습의 한 형태 [1]. 전통적인 단일 에이전트 강화학습에서는 하나의 에이전트만이 환경과 상호 작용하여 최적의 정책을 학습한다. 하지만 실세계 문제 해결 상황에서는 다수의 에이전트가 서로 소통하며 문제를 해결해야 하는 경우가 더 흔하기 때문에, 다중 에이전트 강화학습 연구는 중요하다. 다중 에이전트 강화학습에서는 각 에이전트가 자신의 행동이 다른 에이전트에 영향을 미칠 수 있음을 인지하고 행동해야 한다.

다중 에이전트 강화학습은 불확실한 환경에서의 확률적이고 연속적인 문제 해결을 위해 실생활과 다양한 연구 분야에서 활용되고 있다. 예를 들어, 다중 무인 항공기(Unmanned Aerial Vehicle, UAV)가 협력하여 감시 시스템을 구축하는 연구에서는, 리더와 팔로워 UAV 간의 소통을 통해 CommNet 기반 감시가 수행되었다 [2]. 또한 스마트 해양 환경에서는 심층 결정 정책 그래디언트(Deep Deterministic Policy Gradient, DDPG)를 활용하여 연합 사물인터넷 네트워크를 구축하고 최적화하는 연구가 진행되었다 [3]. 강화학습 방식에 인간 개입을 추가하여, 강화학습을 통해 드론이 목적지를 찾아가도록 훈련시키고, 환경의 난이도가 높을 시 필요하다면 인간의 개입을 통해 훈련된 에이전트를 보정하는 연구도 수행되었다 [4]. 본 논문에서는 특히 다수의 에이전트가 협업을 진행할 때 자율적으로 그룹을 형성하고 행동을 수행하는 방식에 대해 구체적으로 다룬다. 또한 가장 널리 활용되는 성능 평가 방식인 StarCraft Multi-Agent Challenge (SMAC)과 Predator Prey 에 대해서 소개한다.

Ⅱ. 자기조직 알고리즘

다중에이전트 강화학습을 적용할 때, 각 에이전트 별로 역할을 부여에 대한 연구가 소개되었다 [5]. 이 접근법은 유사한 역할을 하는 에이전트들을 그룹화하여, 각 그룹이 유사한 정책과 책임을 가지도록 하는 것을 포함한다. 이를 위해, 에이전트의 정책을 특정 조건에 따라 조절하고, 행동을 통해 역할을 식별하며,



[그림 1] 3-button 문제와 conductor 선출

하위 태스크에 특화된 두 가지 정규화 메커니즘을 도입하였다. 또한 리더를 선정하고 메시지 요약을 통해 통신량을 줄이는 방법에 대한 연구가 수행되었다 [6]. 그림 1 에서처럼 3 개의 버튼을 동시에 누르는 문제를 해결해야 할 때, [그림 1-(a)]에서 버튼 2 를 누르는 에이전트는 누군가 버튼 1 을 누르고 있는지를 알 수 없다. 반면, [그림 1-(b)]에서처럼 에이전트 C 를 conductor 로 삼으면, C 는 에이전트 A, B 와 소통하면서 환경 전체의 상황을 공유할 수 있게 된다. 일정 time step 마다 일부 에이전트가 conductor 로 선출된다. Conductor 를 선출하는 방식에는 각각 일정하게 정해진 확률로 선출하는 random 방식, 그룹 내에 최대한 다양한 conductor 가 선출되도록 하여 일반화 능력을 향상시킨 determinantal point process (DPP) 방식, 그리고 conductor election 문제를 하나의 강화학습 문제로 생각하여 보상을 최대화시키는 관점으로 접근한 policy gradient (PG) 방식을 제안하였다. 또한 Message summary 를 적용하여, 필수적인 정보는 보존하면서 비필수 정보를 제거함으로써 통신 대역폭을 크게 줄이는 방법을 함께 도입하였다. 또한 전체 환경을 관찰하는 'coach'와 일부 환경만을 관찰하는 'player' 간의 협업을 통한 학습 방법을 제안하였다 [7]. 여기서 coach 는 전략 벡터를 생성하고 이를 다른 에이전트들에게 분배하며, player 는 이를 받아 적절한 팀을 구성한다. 이후 coach 는 결과에 따라 정책을 업데이트한다. 이러한 방식으로 훈련된 에이전트들은 zero-shot generalization 을 통해 전혀 다른 팀 구성에서도 효과적으로 작동할 수 있다는 것이 입증된다.

Ⅲ. 성능 평가 방식

[그림 2] SMAC 환경에서의 실험 예시 (8m)

제안하는 알고리즘의 성능 평가를 위해 사용되는 다양한 tool 들이 있다. 다중 에이전트 강화학습의 성능 분석에 가장 널리 활용되는 도구는 SMAC 이다 [8]. SMAC 은 다중 에이전트 강화학습 연구를 위해 20 가지 이상의 다양한 전투 시나리오를 제공한다. 이 시나리오들은 n 개의 에이전트와 m 개의 적 유닛으로 구성되며, 각 유닛은 ID, 이동 능력, 체력, 방어력 등의 특성을 feature 로 가진다. [그림 2]는 이러한 학습 환경의 예시로, 8 기의 Marine 유닛이 동일한 적과 대치하는 시나리오를 보여준다. 이 환경에서 에이전트들은 다른 에이전트의 특성을 부분적으로 관측할 수 있다. 더욱 발전된 형태인 SMACv2 는 기존 SMAC 의 기능에 추가하여, 유닛들의 시작 지점과 유닛 종류를 무작위화하고, 각 유닛의 시야와 공격 범위를 그들의 특성에 맞게 조정할 수 있도록 했다 [9]. 이러한 업데이트를 통해 기존의 다중 에이전트 강화학습(MARL) 알고리즘 실험에 비해 랜덤성을 증가시키고, 다양한 상황에서 알고리즘의 성능을 테스트할 수 있는 환경을 제공한다. 이와 같은 도구들의 사용은 연구자들이 다양한 환경과 시나리오에서 알고리즘을 실험하고 평가할 수 있게 해, 알고리즘의 범용성과 효율성을 검증하는 데 핵심적인 역할을 한다.

Predator Prey [10] 는 여러 Predator 에이전트가 협력하여 Prey 를 사냥하는 가상의 환경이다. 2D 환경에서 Predator 는 상하좌우 방향으로의 이동과 정지까지 5 개의 행동을 수행하며, Prey 는 환경 설정에 따라 동일한 방식으로 이동하거나 정지하도록 한다. 최종적으로 여러 Predator 가 협력하여 최단 시간에 Prey 를 사냥하는 것을 목표로 한다. Predator Prey 는 사용자가 제안하는 알고리즘에 맞게 환경을 가정하기 용이하고 단순하여 구현이 쉽기 때문에 다중에이전트 강화학습의 성능 평가에 널리 활용되고 있다.

Ⅳ. 결론

본 논문에서는 자기조직을 위한 심층강화학습 및 다중에이전트 강화학습의 평가 방식에 대해 다루었다. 다중 에이전트 강화학습의 핵심 요소인 에이전트에 역할을 부여하는 방식, 효율적인 통신을 위한 Conductor 의 선출 방식, 그리고 서로 다른 관측 범위를 가진 에이전트의 적용 방법을 소개하였다. 더불어, 이러한 방법들이 적용된 연구에서 널리 사용되는 성능 평가 도구인 StarCraft Multi-Agent Challenge (SMAC)과 Predator Prey 에 대해서도 설명하였다.

실제 환경에서는 에이전트 간의 협력이 매우 중요한 역할을 하므로, SMAC 등의 평가 도구를 통해 알고리즘을 훈련시킨 후 실제 환경에서의 테스트가 필수적이다. 이는 알고리즘의 실질적인 적용성과 효율성을 검증하는 데 중요한 단계이다. 본 논문은 이러한 방법론과 도구들을 통해 다중 에이전트 강화학습 분야에서의 발전을 촉진하고, 더 넓은 범위의 실제 세계 문제 해결에 기여할 수 있는 기초를 제공하고자 한다.

ACKNOWLEDGMENT

이 성과는 2023 년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF-2022R1A2C2004869). 본 논문의 교신저자는 김중헌임.

참 고 문 헌

[1] A. Dorri, S. S. Kanhere and R. Jurdak, "Multi-Agent Systems: A Survey," *IEEE Access*, vol. 6, pp. 28573-28593, Apr. 2018

[2] W. J. Yun et al., "Cooperative Multiagent Deep Reinforcement Learning for Reliable Surveillance via Autonomous Multi-UAV Control," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 10, pp. 7086-7096, Oct. 2022

[3] D. Kwon, J. Jeon, S. Park, J. Kim and S. Cho, "Multiagent DDPG-Based Deep Learning for Smart Ocean Federated Learning IoT Networks," *IEEE Internet of Things Journal*, vol. 7, no. 10, pp. 9895-9903, October 2020

[4] H. Lee and S. Park, "Sensing-Aware Deep Reinforcement Learning With HCI-Based Human-in-the-Loop Feedback for Autonomous Nonlinear Drone Mobility Control," *IEEE Access*, vol. 12, pp. 1727-1736, Jan. 2024

[5] T. Wang, et al., "ROMA: Multi-agent reinforcement learning with emergent roles,", in *Proc. International Conference on Machine Learning (ICML)*, Jul. 2020, pp. 9876-9886.

[6] J. Shao, et al., "Self-Organized Group for Cooperative Multi-agent Reinforcement Learning, Source models for VBR broadcast-video traffic," in *Proc. Conference on Neural Information Processing Systems (NeurIPS)*, pp. 664-671, December 2022

[7] B. Liu et al., "Coach-Player Multi-agent Reinforcement Learning for Dynamic Team Composition," in *Proc. International Conference on Machine Learning (ICML)*, Jul. 2020

[8] M. Samvelyan, et al., "The StarCraft Multi-Agent Challenge", *arXiv preprint, arXiv: 1902.04043*, Feb. 2019

[9] B. Ellis et al., "SMACv2: An Improved Benchmark for Cooperative Multi-Agent Reinforcement Learning," *arXiv preprint, arXiv: 2212.07489*, Dec. 2022

[10] P. Stone and M. Veloso, "Multiagent systems: A survey from a Machine Learning Perspective," *Autonomous Robots*, vol. 8, pp. 345-383, June 2000