연합 강화 학습에서 가중치 기반 성능 가속화

임현교*, 허주성* 최호빈**, 한연희**[1]
*한국기술교육대학교 창의융합공학협동과정
**한국기술교육대학교 컴퓨터공학과

{glenn89, chil1207, chb3350, yhhan}@koreatech.ac.kr

Weight-based Learning Performance Accelerating in Federated Reinforcement Learning

Hyun-Kyo Lim*, Ho-Bin Choi**, Youn-Hee Han**
*Interdisciplinary Program in Creative Engineering
**Department of Computer Science and Engineering
Korea University of Technology and Education

요 약

최근에 강화 학습 분야에서는 학습의 성능을 높이기 위한 방안으로 다중 에이전트를 활용한 알고리즘 및 시스템들이 활발히 연구되고 있다. 본 논문에서는 기존에 연구한 연합 강화 학습을 가속화 할 수 있는 방안을 제시한다. 이를 위하여 각 에이전트들의 학습 정도를 고려하여 가중치를 정하고 해당 가중치에 따라 Gradient Sharing 을 수행하고, 학습의 촉진을 위해 기존의 연구에서의 단순한 Transfer Learning 이 아닌 현재의 에이전트의 학습의 정도를 고려한 Transfer Learning 을 수행한다. 확장된 연합 강화 학습의 검증을 위해 OpenAI Gym 의 시뮬레이션 환경을 이용해 검증한다.

Ⅰ. 서 론

최근 강화 학습(Reinforcement Learning) 기술이 발전하고 다양한 분야에 적용되고 있으며, 특히 로보틱스 분야에서 강화 학습은 대표적인 응용분야로 로봇의 자동 제어를 위해 사용되고 있다. 기존에는 PID 제어기(Proportional-Integral-Differential controller)를 사용해 실험적/경험적으로 반복하여 적절한 수학적 수식과 제어 파라미터를 찾은 후 로봇을 제어한다. 그러나 강화 학습을 적용할 경우 각 로봇의 수학적 수식을 생성하지 않고 자동으로 강화 학습 에이전트가 지속적인 학습을 통해 최적의 제어가 가능하다. 또한, 멀티-에이전트 기반의 강화 학습을 이용해 다수의 기계를 동시에 훈련 시킴으로써 학습의 성능과 속도를 높이기 위한 연구가 활발히 이루어지고 있다 [1].

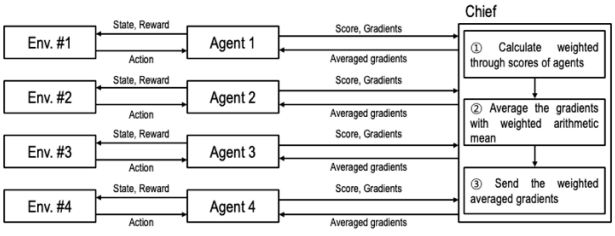
다수의 에이전트들을 동시에 학습 시킬 수 있는 방법으로 연합 정책(Federation Policy)을 적용한 연합 강화 학습 방법이 최근 많이 사용되고 있다 [2]. 연합 정책은 구글에서 2017 년 제안했으며, 분산된 다수의 에이전트들이 학습 모델을 클라우드로 전송함으로써 서로의 학습 경험을 공유하고, 또한 개인정보보호 문제를 해결할 수 있다.

본 논문에서는 기존의 연구[3]를 확장하여 연합 강화 학습(Federated Reinforcement Learning)의 성능과 속도를 증가 시키는 방법에 대해 제안한다. 이를 위하여, 기존에 제안한 연합 정책인 Gradient Sharing 기법과 Transfer Learning 기법을 개선한다. 또한, 강화 학습 알고리즘에서 최근 좋은 성능을 보이고 있는 Actor-Critic 기반 PPO (Proximal Policy Optimization) 알고리즘을 사용한다. 마지막으로 제안하는 연합 강화 학습을 OpenAI Gym 의 CartPole, MountainCar-Continuous, Acrobot, Pendulum 4 가지의 시뮬레이션 환경에 적용해 성능을 검증한다.

Ⅱ. 새로운 연합 정책

본 논문에서 제안하는 연합 강화 학습의 연합 정책은 이전의 연합 강화 학습 방법들과 두가지의 다른 점이 있다. 기존의 연합 강화 학습에서는 학습이 완료된 이후에 경험을 공유하는 방법이다. 그러나 본 논문에서 제안하는 새로운 연합 정책은 학습이 완료된 이후에 학습 모델을 교환하여 학습의 경험을 공유하는 것 뿐만 아니라, 학습을 진행과 동시에 에이전트의 학습 파라미터를 공유하여 학습의 속도를 촉진 시킨다. 이를 위하여 기존의 연구[3]에 확장하여 Weight-based Gradient Sharing 과 Weight-based Transfer Learning 기법을 사용한다.

i. Weight-based Gradient Sharing



(그림 1) Weight-based Gradient Sharing 기법

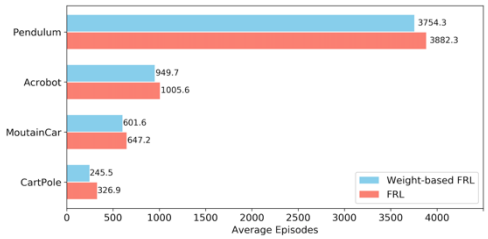
---

[1] 교신저자: 한연희

이 논문은 2018 년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No. 2018R1A6A1A03025526 및 No. NRF-2020R1A6A3A13073735).

(표 1) Chief 에서 Weight-based Gradient Sharing 계산

	Agent 1	Agent 2	Agent 3	Agent 4
gradient	g_1	g_2	g_3	g_4
score	s_1	s_2	s_3	s_4
weighted	$w_1 = \frac{s_1}{s_1+s_2+s_3+s_4}$	$w_2 = \frac{s_2}{s_1+s_2+s_3+s_4}$	$w_3 = \frac{s_3}{s_1+s_2+s_3+s_4}$	$w_4 = \frac{s_4}{s_1+s_2+s_3+s_4}$
gradient	$G_{average} = w_1 \times g_1 + w_2 \times g_2 + w_3 \times g_3 + w_4 \times g_4$			



(그림 2) 확장된 연합 강화 학습 성능 비교

표 2. 현재 학습 정도에 따른 가중치를 적용한 Weight-based Transfer Learning 계산

	Agent 1	Agent 2	Agent 3	Agent 4
current parameter	θ_1	θ_2	θ_3	θ_4
score	s_1	s_2	s_3	s_4
weighted	$w_1 = \frac{s_1}{Terminal\_condition}$	$w_2 = \frac{s_2}{Terminal\_condition}$	$w_3 = \frac{s_3}{Terminal\_condition}$	$w_4 = \frac{s_4}{Terminal\_condition}$
weighted transferred parameter	$\theta_1 = w_1 \times \theta_1 + (1-w_1) \times \theta_{complete}$	$\theta_2 = w_2 \times \theta_2 + (1-w_2) \times \theta_{complete}$	$\theta_3 = w_3 \times \theta_3 + (1-w_3) \times \theta_{complete}$	$\theta_4 = w_4 \times \theta_4 + (1-w_4) \times \theta_{complete}$

다수의 에이전트들은 동일한 목적을 갖는 환경들을 독립적으로 가지고 있으며, 에이전트들은 각각의 환경과 통신을 통해 강화 학습을 수행하게 된다. 에이전트는 학습을 통해 Gradient 를 생성하고 이를 중앙의 Chief 서버에 전달한다. Chief 서버는 각 에이전트들로부터 받은 Gradient 를 일반적인 평균으로 계산하여 공유하는 것이 아니라, 각 에이전트의 현재의 학습 정도를 고려하여 가중치를 계산하고 해당 가중치에 따라 비율을 정하여 각 에이전트의 학습 모델의 Gradient 에 가중치 산술 평균(Weighted Arithmetic Mean)을 적용하여 계산된 Gradient 를 각 에이전트들에게 전달한 후 학습을 진행한다. 그림 1 은 Weight-based Gradient Sharing 기법의 전반적인 모습을 나타낸 그림이다. 표 1 은 Chief 에서 각 에이전트의 현재 학습의 정도를 나타내는 점수를 이용하여 가중치를 계산하고, 각 에이전트들의 Gradient 에 산술 평균을 적용하여 각 에이전트들에게 가중치 산술 평균이 적용된 $G_{average}$ 를 보낸다.

ii. Weight-based Transfer Learning

본 논문에서 제안하는 Weight-based Transfer Learning 기법은 어느 한 에이전트에서 학습이 완료된 경우, 해당 에이전트의 완료된 학습 모델 파라미터($\theta_{complete}$)를 나머지 에이전트들에게 전이 시킴으로써 다른 에이전트들의 강화 학습의 속도를 촉진 시키는 방법이다.

단순히 학습 모델 파라미터 만을 전이 하는 것이 아니라 현재 학습을 수행하는 나머지 에이전트의 학습 수준을 고려하여 전이하게 된다. 이를 위해, Chief 에서는 어느 한 에이전트가 학습이 완료 되면, 해당 에이전트의 완료한 학습 모델 파라미터와 나머지 에이전트들의 현재 점수를 받는다. 점수들을 이용해 표 2 에서 보이는 것처럼 가중치를 계산한다($Terminal\_condition$ 은 각 환경의 학습 종료 조건). 계산된 가중치를 고려하여 학습이 잘 되고 있는 에이전트의 경우, 현재의 파라미터에 더 가중치를 준다. 그러나 학습이 잘 되지 않는 에이전트의 경우 학습이 완료된 에이전트의 학습 파라미터에 더 높은 가중치를 주어 전이를 하게 된다.

Weight-based Transfer Learning 기법은 현재 에이전트의 학습의 수준을 고려하여 전이를 수행하기 때문에 달성하고자 하는 목적은 갖더라도 실제 물리적 특성에 차이가 존재하는 실제 디바이스나 로봇의 경우 기존의 학습 모델의 학습 정도를 고려해 주어야 하기 때문에 전반적인 학습이 안정적이고 촉진 된다.

Ⅲ. 실험 및 검증

본 논문에서 제안하는 확장된 연합 강화 학습의 효과를 검증하기 위해 OpenAI Gym 의 CartPole, MountainCar-Continuous, Acrobot, Pendulum 을 이용한다. 본 연구는 실제 디바이스에 적용하기에 앞서 시뮬레이션 환경에서 검증을 위한 것으로, 실제 디바이스 환경과 동일하게 구성 하기 위하여 동일한 디바이스 장치일지라도 물리적/동적 특성의 차이가 존재하는 것을 반영해 각 시뮬레이션 환경에 작은 노이즈를 추가했다. 또한, 연합을 위하여 각 시뮬레이션 환경 마다 10 번씩 반복적으로 실험을 했으며, 연합을 위하여 4 개의 에이전트들을 사용했다.

그림 2 는 기존의 연합 강화 학습 방법과 확장된 연합 강화 학습을 4 가지 시뮬레이션 환경을 통해 비교한 그래프 이다. 실험은 시뮬레이션 환경 마다 평균적으로 몇 에피소드에 4 개의 에이전트들 모두가 학습이 완료되는 지를 나타내는 그래프이다. 전반적으로 본 논문에서 제안하는 확장된 연합 강화 학습의 성능이 기존의 연합 강화 학습 보다 빠르게 학습이 종료 되는 것을 볼 수 있다.

Ⅳ. 결론

본 논문에서는 기존의 연합 강화 학습의 성능을 높일 수 있는 확장된 연합 강화 학습을 제안한다. 제안하는 연합 강화 학습은 Weight-based Gradient Sharing 과 Weight-based Transfer Learning 을 이용했으며, OpenAI Gym 의 시뮬레이션 환경을 통해 유효성을 검증했다.

참 고 문 헌

[1] Z. Kaiqing et al., “Multi-Agent Reinforcement Learning: A Selective Overview of Theories and Algorithms.” ArXiv abs/1911.10635, 2019.

[2] C. Nadiger, et al., "Federated Reinforcement Learning for Fast Personalization," 2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE), Sardinia, Italy, pp. 123-127, 2019.

[3] H.K. Lim, et al., “Federated Reinforcement Learning for Training Control Policies on Multiple IoT Devices,” Sensors, 20(5), 2020.