휴먼 플레이 데이터를 사용한 '스페이스 인베이더' 게임의 모방학습

이승현[O1] 김경중[2]
[1]용인한국외국어대학교부설고등학교
[2]세종대학교 컴퓨터공학과
andyleewe16@gmail.com, kimkj@sejong.ac.kr

Imitation Learning of 'Space Invaders' using Human Play Data

Andrew Lee[O] Kyung-Joong Kim
[1]Hankuk Academy of Foreign Studies
[2]Department of Computer Science and Engineering, Sejong University

요 약

모방학습(Imitation Learning)은 에이전트가 최상의 성능을 얻기 위해 전문가의 행동을 모방하려고 하는 순차적인 일련의 학습방법이다. 전문가의 정책을 그대로 따라하는 것부터 전문가의 정책으로부터 더 발전해 나갈 수도 있으며, 현재 머신러닝에서 강화학습으로 잘 해결되지 않는 문제들은 모방학습을 통해 해결하려고 한다. 본 논문에서는 MFEC(Model-Free Episodic Control) 에이전트가 직접 학습하여 수집한 샘플 데이터들을 제외하고, 휴먼 플레이 데이터 샘플들을 이용하여 사전 갱신한 모델을 모방학습 실험에 사용하였다. 모방학습을 통해 학습시킨 에이전트가 그렇지 않은 에이전트보다 각 프레임 수에서 더 높은 점수를 얻었고, 빨라진 학습속도로 모방학습의 효용성을 확인하였다.

1. 서 론

강화학습(Reinforcement learning)이란 기계학습의 한 영역이다[1]. 주어진 환경 안에서 에이전트가 현재의 상태를 환경으로부터 인식하여 가능한 액션을 하였을 경우 환경으로부터 주어지는 보상을 최대화 하는 정책을 찾아간다. 주어진 환경은 Markov Decision Process(MDP)로 정의하는데, 이는 주어진 환경을 수학적으로 정의한 것이다. MDP에 따르면 환경은 상태(State), 행동(Action), 상태 전이율(State transition probability), 보상(Reward), 할인률(Discount factor)의 5가지 구성요소로 정의한다. 그림 1과 같이 에이전트는 환경 안에서 환경과 상호작용을 통해 최적의 정책(Policy)를 찾아내는 것이 목적이다. 정책은 확률로 표현할 수 있으며, 특정 상태에서 실행 가능한 행동들의 확률이다(예를 들면, 네 개의 행동이 있다). 가능할 경우 25% 35% 15% 25%의 형태로 표현할 수 있다. 이 강화학습을 이용해서 Google Deepmind는 성공적으로 Atari사의 몇몇 게임들에서 높은 점수를 얻는데 성공했지만, '몬테주마의 복수'를 비롯한 몇몇 게임에서는 매우 낮은 점수를 기록했다[2].

엘론 머스크의 인공지능 연구재단 OpenAI에서 강화학습을 쉽게 학습하고 적용해 볼 수 있는 플랫폼 Gym(https://gym.openai.com)을 출시하였다. 다양한 알고리즘들을 손쉽게 비교할 수 있으며, 학습을 위해 필요한 환경을 구축하는데 드는 시간과 비용을 절감할 수 있다.

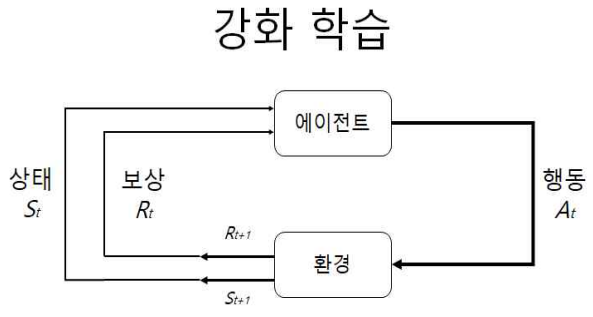

그림 1 강화학습 알고리즘' 의 이해를 위한 도해

OpenAI Gym에는 대표적으로 다양한 Atari 게임들이 있으며, 그 외에도 MuJoCo, Robotics, Toy Text, Box2D 등의 환경에서도 손쉽게 강화학습을 테스트해 볼 수 있는 환경을 제공하고 있다(그림 2).

모방학습(Imitation Learning)은 에이전트가 최상의 성능을 얻기 위해 전문가의 행동을 모방하려고 하는 순차적인 일련의 학습방법이다[3]. 전문가의 정책을 그대로 따라하는 것부터 전문가의 정책으로부터 더 발전해 나갈

수도 있으며, 현재 머신러닝에서 강화학습으로 잘 해결되지 않는 문제들은 모방학습을 통해 해결하려고 시도하고 있다. 강화학습에서는 에이전트가 학습에 필요한 학습 샘플 데이터들을 직접 수집하지만, 모방학습에서는 주어진 전문가의 샘플 데이터들이 있다. 이 샘플 데이터로부터 정책 발전과정을 통해 에이전트의 초기 정책을 개선한다[3].

그림 2 Open AI Gym에서 실행 가능한 Space Invaders

2. 휴먼 플레이 데이터를 이용한 모방학습 제안

MFEC(Model-Free Episodic Control)는 일반적으로 강화학습에서 많이 사용하는 신경망 모델이 아닌 테이블기반 학습 방법이다[4]. 메모리기반 학습 방법의 한 종류이며, 사람이 기억을 토대로 자주 마주치는 환경에서 문제를 해결하는 방식을 모델링한다. 각 상태, 행동 쌍은 Q^{EC} 테이블에 저장하며, 메모리상에도 저장한다. MFEC는 신경망의 느린 학습속도를 대신할 수 있으며, Q^{EC} *(s, a)* 테이블을 통해 강화학습의 함수를 유추한다[4]. Q^{EC} *(s, a)* 테이블은 개별 에피소드가 끝날 때 마다 업데이트가 이루어지며, 상태의 입력으로 가능한 행동을 평가할 수 있다. MFEC는 탐욕적인(greedy) 방식으로 업데이트가 이루어지는데 이것은 비결정적 환경에서의 빠른 학습속도를 보장한다.

본 논문에서는 MFEC 에이전트가 직접 학습하여 수집한 샘플 데이터들을 제외하고, 휴먼 플레이 데이터 샘플들을 이용하여 사전 갱신한 모델을 실험에 사용하였다. 기존 방법보다 낮은 탐험률을 학습에 사용하며, 이는 데이터셋으로부터 더 많은 참조를 통해 전문가 액션의 비중을 늘려 빠른 학습을 가능하게 한다.

3. 결과 분석

에이전트를 학습시킬 플레이 데이터를 모으기 위해서 스페이스 인베이더를 4주간 매일 4시간동안 플레이해서 10만개 이상의 플레이 데이터를 모았고, 이를 에이전트에 학습시키기 시작했다. 모방학습을 시킨 에이전트와 모방학습을 시키지 않은 대조군을 모두 100만 프레임 까지 학습시켰으며, 학습 과정에서 랜덤 액션을 배제한 플레이로 얻은 점수들을 만 프레임마다 기록하였다.

그림 3에서 볼 수 있듯이, 모방학습을 통해 학습시킨 에이전트가 그렇지 않은 에이전트보다 각 프레임 수에서 더 높은 점수를 얻고 있는 것을 확인할 수 있었고, 모방학습의 효용성을 볼 수 있었다.

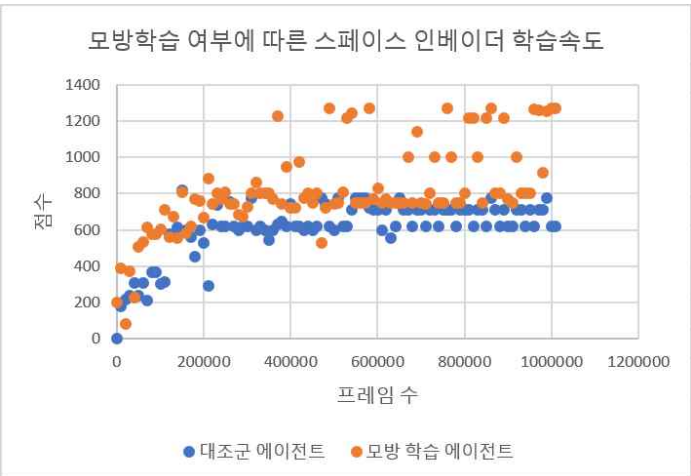

그림 3 모방학습 여부에 따른
스페이스 인베이더 게임 학습속도

4. 결론 및 향후 연구

본 연구를 통해 인간이 직접 게임을 한 샘플 데이터를 수집하여 이를 학습시키면 MFEC 에이전트가 훨씬 더 빠르게 게임을 학습한다는 사실을 실험으로 확인할 수 있었다. 향후에는 모방학습을 이용해 각 Atari 게임에 맞는 맞춤 AI 모델을 제작할 예정이다.

참 고 문 헌

[1] Oh, IS., Cho, CH., Kim KJ. Playing real-time strategy games by imitating human players' micromanagement skills based on spatial analysis. Expert systems with applications, 71, 192-205, 2017

[2] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Hassabis, D. Human-level control through deep reinforcement learning. Nature, 518, 529-533, 2015

[3] Attia, A., Dayan, S. Global overview of imitation learning. arXiv:1801.06503, 2018

[4] Blundell, C., Uria, B., Pritzel, A., Li, Y., Ruderman, A., Leibo, J. Z., Rae, J., Wierstra, D., Hassabis, D. Model-free episodic control, arXiv:1606.04460, 2016