

Baba is Gym : 게임 추론 및 이해에 대한 새로운 도전적인 환경 제안

박태화^o 정의진 김경중^{*}

광주과학기술원 융합기술학제학부

세종대학교 건설환경공학과

광주과학기술원 융합기술학제학부

taehwa-p@gist.ac.kr jin.prelude@gmail.com kjkim@gist.ac.kr

Baba is Gym: A New Challenge for Game Reasoning and Understanding

Taehwa Park^o Euijin Jeong Kyung-Joong Kim^{*}

School of Integrated Technology, GIST

Sejong University Civil and Environmental Engineering

School of Integrated Technology, GIST

요 약

아타리나 스타크래프트 II와 같은 게임 환경들은 심층 강화학습 알고리즘의 성능을 측정하기 위하여 자주 사용되어왔다. 그러나, 이런 환경들은, 환경에서 제공하는 객체들의 깊은 이해 없이 게임을 푸는 것이 가능하였다. 본 논문에서는 환경의 이해 없이 풀기 어려운 'baba is gym' 환경을 제안한다. 환경의 구조와 규칙을 설명함으로 본 환경이 왜 강화학습으로 풀기 어려운지 설명하며, 심층 강화학습 알고리즘의 새로운 도전과제를 제안하도록 한다.

1. 서 론

아타리 환경[1]은 Deep Q-Learning(DQN)[2] 기법의 성능을 측정하기 위해 사용한 플랫폼이다. 이 플랫폼은 심층 강화학습에서 성능을 측정하는 환경으로 사용되고 있다. 스타크래프트 II와 같은 환경[3]도 멀티에이전트 문제에서 사용되고 있다.

위의 환경들은 게임 인공지능 분야에서 자주 사용되고 있는 환경이지만, 새로운 게임 환경에 대한 요구가 없는 것은 아니다. 그 이유 중 하나는 심층 강화학습 기법으로 학습된 인공 신경망이 객체에 대한 이해가 부족하면 게임을 못 푸는 모습도 보이기 때문이다. 예를 들어 PacMan은 매 스테이지를 풀 때마다 스테이지의 모양과 색이 변화한다. 인공 신경망은 입력의 변화에 매우 민감한데, 게임의 규칙은 변하지 않았더라도 관측되는 상태가 변했기 때문에 인공 신경망은 변화하는 스테이지에서 좋은 성능을 내지 못하는 모습을 보인다.

사람의 경우 이런 문제는 발생하지 않는다. 이는 사람이 게임의 규칙을 이해하고 게임에서 제공하는 객체의 의미를 이해하고 있기 때문이다. 비록 인공 신경망이 입력의 변화에 민감하게 반응할지라도, 객체에 대한 이해는 사람보다 떨어진다. 이러한 과제를 해결하기 위해 게임의 규칙 이해와 객체의 추론 능력이 중요한 새로운 게임 환경이 필요하다.

본 논문에서는 'baba is gym'(BIGym)이라는 환경을 제안한다. 본 환경은 2차원 퍼즐게임 'baba is you'(BIY)

[4]의 미니게임으로 미로에서 지정된 위치로 상자를 옮기는 게임인 Sokoban과 유사해 보이지만, 객체의 조합으로 규칙을 생성하는 특징 때문에 객체의 이해 없이 풀기 매우 어려운 게임이다.

2. 배 경

BIY는 게임 개발자 Arvi Teikari[5]가 2017년 Nordic Game Jam contest에서 처음 공개하였으며, 2019년 3월 13일 공식적으로 판매를 시작했다. 본 게임에서는 200개 이상의 스테이지와 121개 이상의 상태와 객체를 제공한다.



그림 1. 객체와 상태의 조합에 따른 다양한 상태

플레이어는 YOU상태를 가진 객체를 움직일 수 있으며, 다음 스테이지로 넘어가기 위해 WIN상태를 가진 객체에 접촉해야 한다. 이외의 규칙은 모두 화면 내의 문자객체 조합으로 표현된다. 이 조합을 바꿈으로 게임의 규칙을 바꿀 수 있으며 스테이지마다 다른

규칙을 조합하여 풀어나간다.

그림 1은 조합에 대한 예시로 (가)는 스테이지의 기본 상태로 벽 객체가 STOP상태를 가지고 있기 때문에 밖으로 이동할 수 없다. 밖으로 이동하기 위해선 (나)와 같이 WALL IS STOP 문자 조합을 깨어 벽 객체가 상태를 갖지 못하게 만든 뒤, 이동을 해야 한다. (다)는 게임의 특징을 보여주는 그림으로 WALL IS FLAG 문자 조합으로 모든 WALL객체를 FLAG객체로 변형한 모습이다. 이러한 특성 때문에 게임의 스테이지를 풀기 위해 객체와 객체 간 관계의 의미를 이해해야 한다.

BIY는 게임을 푸는데 객체와 관계에 대한 깊은 이해를 요구하지만, 게임을 하는 경우 상, 하, 좌, 우, 정지 5개의 deterministic action space만을 요구한다. 이는 DQN과 같은 이산적인 행동을 출력하는 심층 강화학습 알고리즘을 적용하기 좋은 환경을 뜻한다. 또한, 게임의 직관적인 특징 덕분에 사람이 직접 여러 개의 스테이지를 생성하는 것이 어렵지 않다. 이 때문에 BIY는 다양한 환경에서 심층 강화학습의 추론 및 인지 능력을 확인하기에 좋은 환경이라 말할 수 있다.

하지만, BIY는 상용게임으로 스타크래프트 II 환경과 달리 심층 강화학습을 위한 API를 제공하고 있지 않다. 스테이지를 사용자가 직접 만드는 것 또한 불가능하다. 이에 본 논문에서는 BIY의 미니게임, Baba is Gym 환경을 제안한다.

3. Baba is Gym 환경

BIGym은 Python 기반 환경으로 pygame 패키지와 Python의 기본 함수만으로 구현되었다. pygame 패키지는 상태를 이미지로 출력하거나 키보드 입력을 받기 위해 사용하며, 환경은 Python 기본 패키지만으로 동작한다.

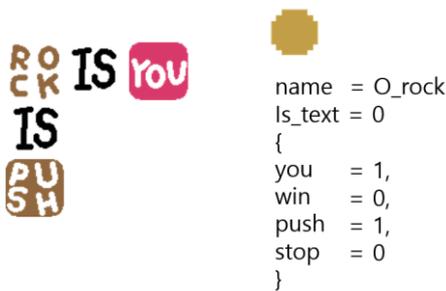


그림 2. BIGym의 규칙 표현 방법과 객체의 상태 표현

BIGym은 BIY와 같은 방법으로 규칙을 표현하며 이는 그림 2와 같이 표현된다. 모든 규칙은 '객체 is 객체' 또는 '객체 is 속성'으로 표현되며 왼쪽에서 오른쪽, 위에서 아래쪽으로 적용된다. '객체 is 속성'의 구조를 가진 규칙은 해당 객체에 속성을 부여하지만, '객체A is 객체B'형태의 규칙은 규칙이 생성된 후 행동마다 모든 객체A를 객체B로 변화시킨다. 이러한 특징을 이용해 '객체A is 객체B is 객체A'라는 규칙을 만들어 매번

움직일 때마다 객체를 변화시키는 복잡한 규칙을 만드는 것도 가능하다.

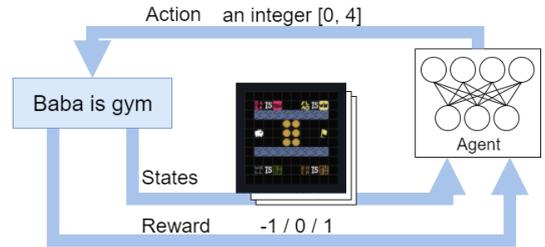


그림 3. 신경망의 학습을 진행할 때의 구조도

본 환경은 심층 강화학습을 연구하는 연구자가 자주 접해본 Open AI Gym[6] 패키지와 유사하게 사용이 가능하다. Gym에서 사용하는 step, reset등과 같은 명령어도 구현되어 있으며 유사한 입력과 출력을 제공한다.

환경의 입력은 0에서 4까지의 정수이다. 각 정수는 0부터 정지, 상, 하, 좌, 우 이동을 나타낸다. 환경의 출력은 다음 상태를 나타내는 이미지 또는 객체집합과 그에 따른 보상이다. 보상은 스테이지를 풀었을 경우 1 스테이지의 진행이 불가능한 경우 -1 두 경우를 제외한 모든 경우에서 0이 주어진다.

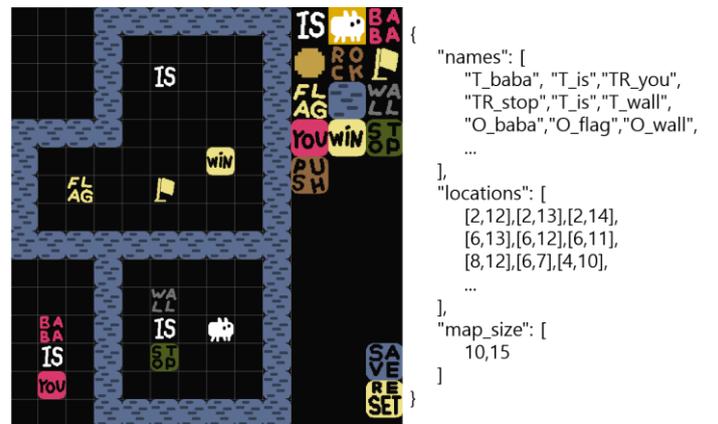


그림 4. 스테이지 에디터(좌)와 생성된 스테이지 파일(우)

BIGym은 매 스테이지의 자유도가 높고 스테이지에서 제공하는 객체에 따라 풀이가 달라지는 특징이 있다. 이 때문에 스테이지를 자동 생성하는데 어려움이 있다. 스테이지를 코드로 구현하면 스테이지를 생성하는데 많은 시간과 노력이 들어간다. 이를 해결하기 위해 스테이지 에디터 또한 구현했다. 스테이지 에디터는 BIGym과 같은 패키지 의존성을 가진다. 에디터는 마우스를 이용한 조작만 가능하며, 배치를 원하는 객체를 선택하고 좌측에 드래그하면 그림판에서 선을 그리듯 객체가 배치된다. 이후 Save버튼을 클릭하여 그림 4의 우측과 같이 json파일로 출력하거나 Reset버튼으로 다른 스테이지를 생성하는 것이 가능하다.

4. 실험 및 결과

본 실험에서는 BIGym이 객체의 이해 없이는 풀기 어려운 문제임을 사람과 무작위 행동을 하는 에이전트, 심층 강화학습 알고리즘의 비교를 통하여 보인다.

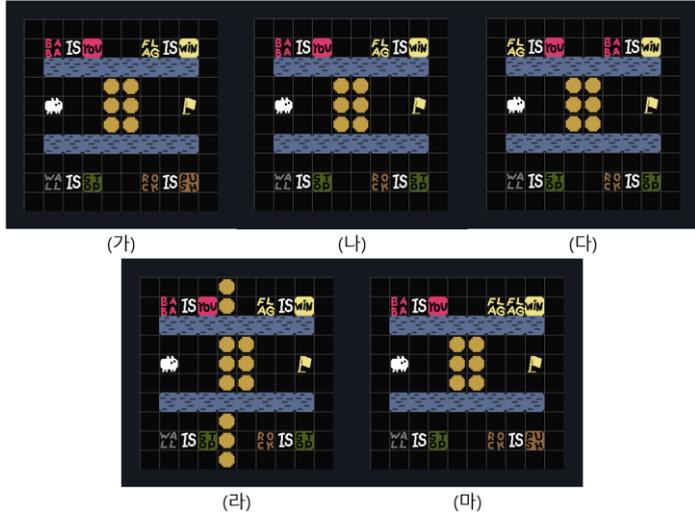


그림 5. 실험에 사용된 5개의 스테이지

실험의 진행을 위하여 생성한 스테이지는 그림 5의 5개와 같다. 이 스테이지들은 매우 비슷해 보이지만, 모든 스테이지를 같은 방식으로 푸는 것은 어렵다. 예를 들어 (가)스테이지와 (나)스테이지는 'Rock is' 뒤에 오는 속성 외엔 달라진 것이 없다. 하지만, 이 변화로 (가)에선 바위를 밀어서 깃발에 닿는 것이 가능했지만, (나)에선 바위를 밀지 못하고 깃발로 돌아서 가야 한다. (가), (나), (다)스테이지의 경우, 탐색을 많이 하면 풀 수 있지만 (라)와 (마)스테이지의 경우, 객체간 관계와 객체를 이해하지 못한다면 풀기 어려울 것이다.

심층 강화학습 알고리즘은 이전 스테이지와 같은 방식으로 풀 수 없다면 잘 풀어나가지 못한다. 이는 보상을 통해 현재 상태와 행동의 관계만을 학습하는 심층 강화학습 알고리즘의 특징 때문이다. 객체를 조합하는 행동은 보상을 주지 않기 때문에 조합이 좋은지 나쁜지 판단하기 어렵다. 이러한 상태에선 그림 5의 (마)스테이지를 깨는 것은 거의 불가능에 가깝다. 더 나아가 스테이지마다 객체를 조합하는 순서가 다르기 때문에 모든 상황에 맞는 보상함수를 설계하는 것도 어려움이 있다.

실험은 심층 강화학습 알고리즘을 학습하여 스테이지를 푸는 횟수와 푸는데 움직인 횟수를 얻어내 이를 비교하는 것으로 진행하였다. 학습엔 DQN 알고리즘을 사용했으며, 알고리즘의 충분한 탐험을 위해 한 스테이지에서 최대 행동 수를 200으로 제한하였다.

비교 대상은 사람과 무작위 행동 모델을 선정하였다. 사람은 다섯 명의 사람이 각각 스테이지를 2번씩 풀어 보는 것으로 진행되었다,

무작위 행동 모델은 랜덤 함수를 이용해 무작위의 행동을 선택하여 10번의 실험을 진행하였다.

모든 모델은 총 10번의 실험을 진행하였으며, 실험을 통해 얻은 성공률과 행동 수의 평균값을 표 1에 기록하

였다.

	스테이지 클리어 확률과 이에 필요한 평균 행동 수 (성공률 / 행동 수)				
	(가)	(나)	(다)	(라)	(마)
사람	100/9	100/17.6	100/17	100/19.2	100/32.1
무작위 행동	10/189.7	0/200	0/200	0/200	0/200
DQN	30/19	10/126	0/200	0/200	0/200

표 1. 각 스테이지에서의 실험 결과

표 1은 DQN 모델의 학습 결과와 사람과 무작위 행동 모델의 성공률과 행동 수를 보여준다. DQN 모델이 무작위 행동보다 나은 결과를 보이지만, 여전히 낮은 성능을 보인다. 이는 보상이 드문 환경일 뿐만 아니라 모델이 (가)스테이지와 (나)스테이지의 차이를 이해하지 못했기 때문이다. 하지만, 사람은 첫 스테이지 이후 이전에 보았던 객체와 규칙을 이해하여 다음 스테이지들을 풀어나가기 때문에 모든 스테이지를 풀 수 있었다.

5. 결론 및 향후 연구

비교적 단순한 게임이지만, 객체의 이해가 게임에서 차지하는 비중이 높은 게임이 연구를 위해 필요하며, 본 논문에서는 "Baba is You" 게임을 연구할 수 있는 환경을 제안하였다. 본 환경은 5개의 객체와 4개의 속성만 지원하지만, 심층 강화학습의 이해능력을 시험할 때 적은 요소만으로도 신경망의 이해능력을 확인할 수 있음을 보였다. 추후 연구로는 신경망의 이해능력을 깊이 있게 시험하기 위하여 객체와 속성을 추가할 예정이다.

감사의 글

본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 출연금 등으로 수행하고 있는 2019년도 문화기술연구개발 지원사업으로 수행되었습니다.

이 논문은 2017년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초 연구 사업임 (2017R1A2B4002164). *교신 저자

참조문헌

- [1] Bellemare, Marc G., et al. "The arcade learning environment: An evaluation platform for general agents." Journal of Artificial Intelligence Research 47 (2013): 253-279.
- [2] Mnih, Volodymyr, et al. "Playing atari with deep reinforcement learning." arXiv preprint arXiv:1312.5602 (2013).
- [3] Vinyals, Oriol, et al. "Starcraft ii: A new challenge for reinforcement learning." arXiv preprint arXiv:1708.04782 (2017).
- [4] Teikari, A. (2019). Baba is You - release date trailer. [online] YouTube. Available at: <https://www.youtube.com/watch?v=U7MJljs0USo> [Accessed 20 Feb. 2019].
- [5] Teikari, A. (2019). Baba Is You. [online] Hempuli.com. Available at: <https://hempuli.com/baba/> [Accessed 14 May 2019].
- [6] Brockman, Greg, et al. "Openai gym." arXiv preprint arXiv:1606.01540 (2016)