ARCLE: 추상화 및 추론을 위한 강화학습 환경*

이호성[O1] 김세진[2] 김선동[2]
[1]광주과학기술원 전기전자컴퓨터공학부 [2]광주과학기술원 AI대학원
gitpush-force@gm.gist.ac.kr, {sejinkim, sundong}@gist.ac.kr

ARCLE: Gymnasium Environment for Abstraction and Reasoning Corpus

Hosung Lee[O1] Sejin Kim[2] Sundong Kim[2]
[1]GIST EECS [2]GIST AI Graduate School

요 약

본 논문에서는 추상화 및 추론 (Abstraction and Reasoning Corpus; ARC) 문제를 위한 강화학습 환경인 ARCLE을 소개한다. ARCLE은 인간의 ARC 문제 해결 기록을 수집하는 웹 인터페이스인 O2ARC에서 영감을 받아 제작되었다. ARCLE은 Gymnasium 프레임워크에서 구현되었으며, ARC의 모든 문제를 풀 수 있는 행동 공간과 상태 공간이 탑재되어 있다. 본 논문에서는 ARCLE의 구조 및 작동 원리, 사용 방법을 상세히 소개한다. ARCLE은 PyPI를 통해 쉽게 설치할 수 있으며 (pip install arcle), 인간의 풀이법을 본떠 문제를 해결해 나가는 강화학습 에이전트를 훈련하는 테스트베드로 활용되고 있다.

1. 서 론

Abstraction and Reasoning Corpus (ARC)[1]는 인공지능의 추상화와 추론 능력을 평가하는 데이터셋이다. ARC의 각 문제에는 3~5 개의 예시 입출력 쌍과 문제 입력 배열이 주어지며, 인공지능은 입출력 관계에서 일정한 규칙을 찾아 이를 문제 입력 배열에 적용하여 출력 배열을 생성해야 한다. 모든 입출력 쌍은 최대 30x30 크기의 배열이며 배열의 요소(픽셀)는 10가지 색깔에 대응되는 0~9의 정수이다. ARC의 각 문항을 해결하기 위해선 물체 구분, 정렬, 셈법, 기초적인 기하 등을 활용해야 한다.

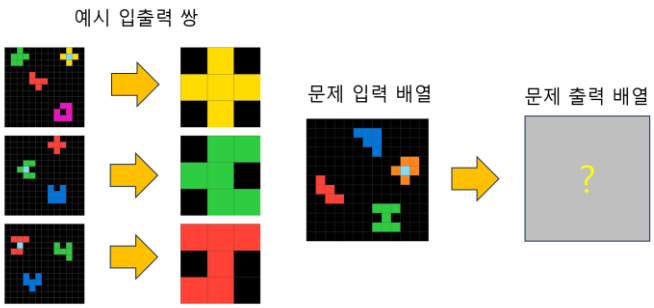

그림 1 ARC 데이터셋의 문제 정의

인간은 그림 1에 정의된 문제를 풀 때 하늘색 픽셀의 위치와 오브젝트를 구분 정보, 오브젝트의 잘라내기를 조합해 쉽게 정답을 도출할 수 있다. 이처럼 인간은 다양한 사전 지식을 조합하여 평균적으로 83.8%의 정확도를 낸다[2]. 그러나, ARC의 모든 문제에 포함된 사전지식을 정리하여 도메인 특화 언어(DSL)를 제작한 후 DSL조합을 탐색하더라도 최대 30%의 정확도를 낸다[3]. 인간에 비해 크게 낮은 정확도는 DSL의 설계 과정이나 DSL 조합 과정 중 적어도 하나에 성능 저하의 요인이 있다는 것을 암시한다.

DSL 조합 과정에 초점을 맞추어 보면, DSL을 합성하는 과정은 강화학습 문제로 생각할 수 있다. 행동 공간은 DSL로, 상태 공간의 초기 상태를 문제 입력 배열로 설정하여 강화학습 환경을 구성할 수 있다. 강화학습 알고리즘을 실행하기 위해선 적절한 강화학습 환경이 주어져야 한다. gym-arc[4] 라는 환경이 존재했으나 오랜 기간 유지보수되지 않았으며, 기본적인 색칠, 크기 조절, 플러드 필만을 사용하여 구현되었다. 본 논문에선 더 많은 DSL 지원을 위해 ARCLE(ARC Learning Environment)을 제안한다. ARCLE은 이동, 회전 등 오브젝트 기반 DSL을 추가한 O2ARC[5] 인터페이스에 착안해 구현되었으며, O2ARC에 정의된 모든 DSL을 구현하는 강화학습 환경이다. 사용자는 사용자 정의 환경을 만들기 위해 자유롭게 행동 및 상태 공간, 보상 메커니즘을 수정할 수 있다. ARCLE은 PyPI를 통해 설치 가능하다 (pip install arcle).

2. ARCLE의 구현

2.1. ARCLE의 기본 구조

ARCLE은 Python 3과 Gymnasium[6] 프레임워크에서 구현한 강화학습 환경 패키지이다. ARCLE의 환경은 마르코프 결정 과정(MDP) 형태이다. 특정 상태에 있는 환경에 에이전트가 행동을 취하면 환경의 상태가 전이하며 에이전트에게 보상이 주어진다. ARCLE의 환경은 에

* 이 논문은 과학기술정보통신부의 재원으로 한국연구재단과 정보통신기획평가원의 지원을 받아 수행된 연구임 (RS-2023-00240062, RS-2023-00216011, 2019-0-01842)

이전트가 편집 중인 배열(상태)에 색을 칠하거나 크기를 바꾸는 등의 행동을 취하는 MDP의 형태를 띤다.

ARCLE의 패키지 구조는 loaders, actions, envs의 세 모듈로 이루어져 있다. 데이터셋을 강화학습 환경에 공급하기 위한 loaders 모듈, 픽셀 이동이나 복사-붙여넣기 등 다양한 행동을 정의한 actions 모듈, 그리고 강화학습 환경을 정의한 envs 모듈이다.

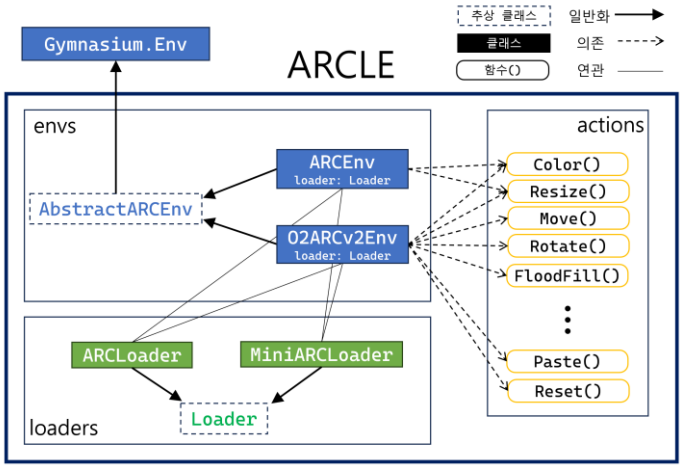

그림 2 ARCLE 패키지 구조. 각 화살표와 도형의 의미는 우측 상단 범례에 표기함.

2.2. loaders 모듈

ARCLE의 loaders 모듈에는 ARCLE의 환경에 ARC 데이터셋을 공급하는 로더가 구현되어 있으며 Mini-ARC[5]와 같이 파생 데이터셋 또한 환경에 공급할 수 있도록 설계되었다. 데이터셋을 공급하기 위해선 ARCLE의 'Loader' 추상 클래스를 상속받고 get_path 메서드와 parse 메서드를 구현하여 사용할수 있다. 현재 ARCLE에는 ARC와 Mini-ARC 데이터셋 로더가 구현되어 있다.

2.3. envs 및 actions 모듈

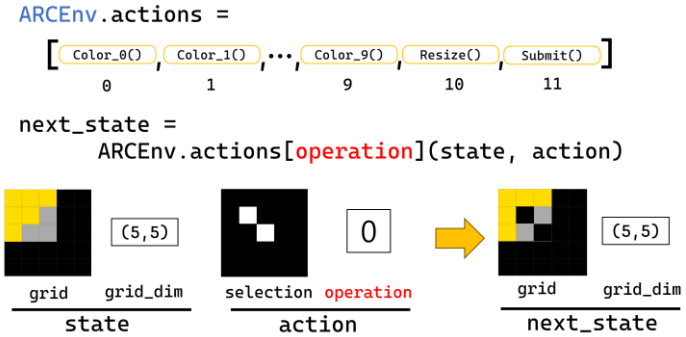

그림 3 ARCEnv의 행동 공간 구조화 방식과 전이 과정에 대한 모식도. 두 픽셀에 0(검은색)을 칠한다.

ARCLE의 envs 모듈에는ARC 강화학습 환경 구현을 위한 'AbstractARCEnv' 추상 클래스가 존재한다. 이 클래스를 상속받아 상태 공간과 행동 공간을 정의하여 새로운 강화학습 환경을 제작할 수 있다. 기본적인 상태 공간은 에이전트가 현재 편집 중인 색깔 행렬 'grid'와 그 행렬의 크기를 나태나는 'grid_dim'으로 구성된다. 행동 공간은 다음과 같이 정의할 수 있다. 우선, actions 모듈에 구현된 다양한 함수를 불러오거나 직접 구현해 리스트로 저장한다. 행동 공간을 함수 리스트의 인덱스 'operation'과 함수의 적용 위치를 나타내는 이진 배열 'selection'으로 구성된다. 그러면 그림 3이 나타내는 것과 같이 'operation'과 'selection'을 이용해 상태가 전이된다.

ARCLE에는 ARCEnv와 O2ARCv2Env, 두 개의 부속 강화학습 환경이 AbstractARCEnv를 상속받아 미리 구현되어 있다. 두 환경 모두 gym.make()로 불러온 후 타 Gymnasium의 환경과 동일한 방식으로 사용할 수 있다.

2.3.1. ARCEnv의 상태 공간 및 행동 공간

ARCEnv는 ARC의 모든 문제를 풀 수 있는 최소한의 상태 공간과 행동 공간이 탑재되어 있다. 상태 공간은 0~9 의 정수로 이루어진 색깔 배열인 'grid', 그리고 배열의 크기를 나타내는 'grid_dim'이 있다. 행동 공간의 'operation'은 그림 3 상단에 표현된 바와 같이 0~11의 정수이다. 0~9는 'selection'으로 선택된 픽셀을 0~9의 색으로 칠하는 것이며, 10은 정답 그리드 크기로 'grid_dim'을 변경, 11은 정답 제출을 의미한다.

2.3.2. O2ARCv2Env의 state와 action 공간

O2ARCv2Env는 사람이 ARC를 푸는 과정을 수집하기 위한 웹 인터페이스인 O2ARC를 보완하여 강화학습 환경 형태로 구현한 것이다. O2ARC는 본래 ARC 데이터셋과 함께 배포된 기본 인터페이스[7]에서 오브젝트 기반 행동을 강화한 것으로, 픽셀을 선택한 이후 이동, 회전, 반전을 자유롭게 할 수 있다. 따라서 행동 공간은 선택된 픽셀들의 위치를 나타내는 이진 배열 'selected'와 이동, 회전, 반전을 보조하기 위한 딕셔너리인 'object_states'를 포함한다. 또한, 복사-붙여넣기 기능을 위한 클립보드 배열인 'clip', 'clip_dim' 역시 포함된다.

행동 공간은 ARCEnv와 유사하게 'selection'과 'operation'으로 이루어져 있으나, 'operation'의 범위가 0~34로 총 35개의 함수가 할당되어 있다. 0~9는 색칠, 10~19는 플러드 필이다. 20~27은 오브젝트 기반 행동으로, 상하좌우 이동, 회전 그리고 반전으로 이루어져 있다. 28과 29는 각각 문제 입력 배열 또는 편집 중인 'grid'로부터 'selection'이 가리키는 픽셀을 클립보드로 복사하며 30은 붙여넣기이다. 31은 문제 입력 배열을 'grid'로 복사, 32는 'grid'를 0(검은색)으로 초기화, 33은 'grid'의 잘라내기, 34는 제출이다. 특히, 20~27번 오브젝트 관련 행동은 선택된 오브젝트가 다른 픽셀을 일시적으로 가리며 지나가는 경우가 발생할 수 있으므로, 'object_states'에 각종 상태를 저장하여 픽셀이 사라지지 않도록 일관성을 유지하였다.

2.4. ARCLE의 시각화

ARCLE은 터미널상에서 시각화된다. ARC의 모든 입력과 출력은 기본적으로 10가지 색상을 이용한 배열로 나타나므로 ANSI Escape Code를 사용하여 그림 4 와

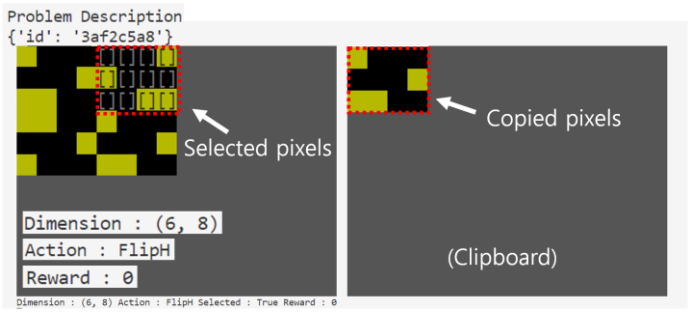

그림 4 O2ARCv2Env의 시각화.

같이 시각화할 수 있다. 터미널 창의 최상단에는 문제 ID, 최하단에는 'grid_dim'과 마지막에 취한 행동, 그리고 마지막으로 얻은 보상을 표시한다. 그림 4는 O2ARCv2Env를 시각화한 것으로, 왼쪽 'grid'와 함께 클립보드를 나타내는 'clip' 배열이 오른쪽에 표시된다. 오브젝트 관련 행동을 취할 때('operation' 20~27)에는 에이전트가 선택한 픽셀들 ('selected') 이 대괄호('[]')로 강조된다. 그림 4의 원본 스크린샷의 글자와 색깔이 작기 때문에 확대하여 표현하였다.

3. O2ARCv2Env의 테스트

ARCEnv에 비해 행동 및 상태 공간 모두 상대적으로 복잡한 O2ARCv2Env는 O2ARC 인터페이스를 통해 사람으로부터 수집한 풀이과정 (ARC를 풀기 위해 사용한 행동 및 편집 중인 색깔 배열의 시퀀스)를 변환하여 환경의 동작을 검증했다. 사람의 풀이과정으로부터 행동을 순차적으로 추출하여 환경에서 실행했을 때 기록된 색깔 배열과 강화학습 환경의 'grid' 및 'grid_dim'이 일치하는지 확인하는 방식으로 동작을 검증하였다. O2ARC 인터페이스에서 수집한 1953개의 기록에 대해 테스트하였고, 풀이과정에 기록된 배열이 환경의 'grid'와 모두 일치함을 확인하였다.

4. 결론 및 향후 연구

4.1. 결론 및 기대 효과

ARCLE은 Gymnasium에 기반한 강화학습 환경 패키지로, 사용자가 고안한 강화학습 알고리즘을 테스트할 수 있는 환경을 제공한다. ARCLE은 ARC 기본 테스팅 인터페이스의 확장인 O2ARC의 상태 공간과 행동 공간을 명확히 구조화하여 확장성을 높였다. 따라서, 사용자의 ARC 해법에 필요한 환경 상태와 행동을 자유롭게 추가 및 제거하여 파생 강화학습 환경을 제작할 수 있다. 궁극적으로는 ARC의 강화학습 해법 연구의 기반으로 ARCLE이 활용되어 강화학습 기반 인공일반지능 개발에 도움이 되기를 기대한다.

ARCLE에 그래픽 인터페이스를 추가하여 사람의 ARC 해결 과정을 수집하는 데에도 사용 가능하다. 수집한 풀이 기록을 사용하여 강화학습 알고리즘을 학습시키거나, 풀이과정에 나타나는 공통적인 행동의 조합을 DSL로 추가하여 인간의 핵심 사전지식을 분석하는 데에 사용 가능하다[2].

4.2. 향후 개선 방향

현재 actions 모듈의 동작 테스트는 O2ARC로부터 얻어진 인간의 ARC 풀이과정과의 일치 여부로 진행하고 있다. 오픈소스 프로젝트로써 다양한 행동을 추가해 나가기 위해 action마다 단위 테스트를 할당하여야 한다. 이는 사용자가 강화학습 알고리즘을 ARCLE에 적용할 때 불필요한 디버깅을 줄여주어 사용자 편의성을 증진한다.

더 나아가, 현재 환경에서 주어지는 보상은 정답 제출 행동 직후의 'grid'와 정답이 완전히 일치할 경우 1, 아니면 0을 반환하는 형태이다. 이러한 방식은 강화학습 에이전트의 보상 총합이 0이 되어 에이전트 학습에 큰 걸림돌이 된다. 이를 방지하기 위해 보상 메커니즘을 개선하여 많은 강화학습 알고리즘이 ARCLE 위에서 성공적으로 최적화될 수 있도록 개선할 것이다.

참 고 문 헌

[1] F. Chollet, "On the Measure of Intelligence," arXiv, preprint arXiv:1911.01547, 2022.

[2] A. Johnson et al., "Fast and flexible: Human program induction in abstract reasoning tasks," arXiv, preprint arXiv:2103.05823, 2021.

[3] Lab42, "Archathon 2022," Lab42, https://lab42.global/past-challenges/arcathon-2022, (accessed Oct. 31, 2023).

[4] MitrofanovDmitry, "gym-arc," GitHub, https://github.com/MitrofanovDmitry/gym-arc, (accessed Oct. 31, 2023).

[5] S. Kim et al., "Playgrounds for Abstraction and Reasoning," NeurIPS nCSI Workshop, 2022.

[6] M. Towers et al., "Gymnasium," Zenodo, 2023. doi:10.5281/zenodo.8127026.

[7] Lab42, "ARCreate Playground," Lab42, https://arc-editor.lab42.global/playground, (accessed Oct. 31, 2023).