Unity ML-Agents를 이용한 사거리 자율주행 환경 구현 및 컴포넌트 설계

박우진○ 이형진 이준성†
경기대학교 기계시스템공학과
(pnltoen@naver.com, e9305@naver.com, jslee1@kyonggi.ac.kr)

Design and Implementation of an Autonomous Driving Environment for Crossroad and Components Using Unity ML-Agents

Woo-Jin Park, Hyeong-Jin Lee, Joon-Seong Lee†
Dept. of Mechanical Engineering, Kyonggi University

요 약

강화학습은 보상을 최대화 하는 방법 중 하나이다. 현재 Unity ML-Agents를 통해 강화학습 진행 및 시각화가 가능하고, 이에 강화학습 기술에 대한 기대가 높아지고 있다. 기존의 공개된 일방향 통행, 에이전트가 서로를 감지 못하는 환경이 아닌 사거리 환경 구축을 통하여 현재 ML-Agents 활용 및 적용가능성을 연구하고자 한다.

1. 서 론

Unity ML-Agents의 강화학습을 통해 구축된 환경에서 설정한 행동을 통해 보상을 최대화함으로써 최적화 문제를 풀 수 있다[1].

현재 공개된 ML-Agents 자율주행 환경은 일방향통행 환경 또는 각각의 에이전트가 서로를 감지(Sensor)하지 못하는 환경이다. 이러한 환경은 데이터간의 상관관계가 낮아 구현 및 결과 예측이 용이하지만 실제 도로상황에 적용하기 어려운 한계점을 갖는다. 에이전트가 서로를 감지하는 사거리 환경의 경우 데이터 상관관계의 복잡도가 높아 학습 및 구현에 어려움을 겪지만 시뮬레이션의 정확도 및 잠재력을 높일 수 있다는 장점을 갖는다.

본 연구에서는 가로 8차선, 세로 10차선의 일반통행 사거리 환경을 구현하고 적절한 보상 및 Epsilon 값을 설계하는 방법을 제시하고자 한다. 또한, 구현 방법과 보상함수 및 Epsilon 값 설계 방법을 토대로 설계단계에서의 의도하지 않은 패턴에 갇히는 문제[2]에 대한 해결방안을 제시하였으며, 구현한 환경을 설명하고, 현재 진행하고 있는 후속 연구 및 방향성을 나타내고자 한다.

2. 환경

2.1 Unity ML-Agentss 선정 이유

많은 메모리 및 리소스를 할당하는 인공지능 학습 특성 상 저사양/소규모 프로젝트에 자주 사용하는 Unity를 선정하였다. 또한 Unity는 환경 구현을 쉽게 할 수 있는 Asset과 Package를 제공하며, 자체 ML Toolkit을 제공하여 환경에 대한 강화학습 알고리즘을 쉽게 적용할 수 있다[1].

2.2 환경 구현 방법

실제와 최대한 유사한 환경을 구현하기 위해 Blender에 BlenderGIS 애드온을 설치하여 강남역 2호선 주변의 건물을 모델링 한 후 그림 1과 같이 구현하였다.

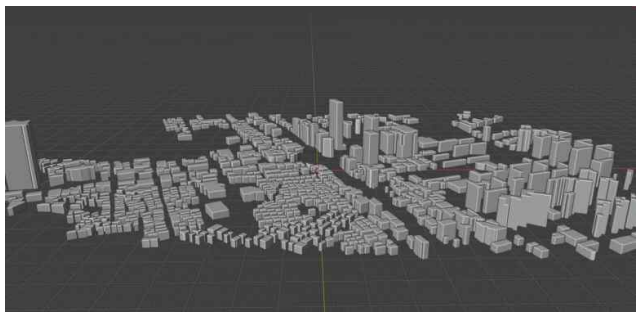


그림 1 구현한 강남역 2호선 주변 건물

이 후 Probuilder(5.0.3) 사용, 강남대로의 차선 길이를 반영하여 3.5m의 도로폭을 그림 2와 같이 구현 하였다.

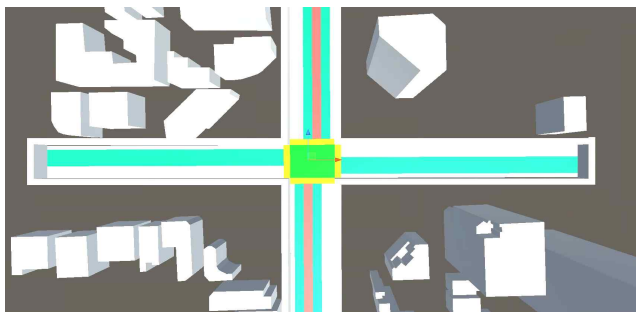


그림 2 차선길이 반영된 도로폭 구현

2.2 체크포인트 구현

에이전트의 역주행 및 차선 침입을 방지하기 위해 Trigger 속성의 체크포인트를 생성하였고 체크포인트 오브젝트의 이름, 태그에 맞춰서 순서를 설정하였다. 체크포인트 스크립트로는 OnTriggerEnter를 사용하였는데 별도의 변수를 선언하여 중복 보상을 방지하였다.

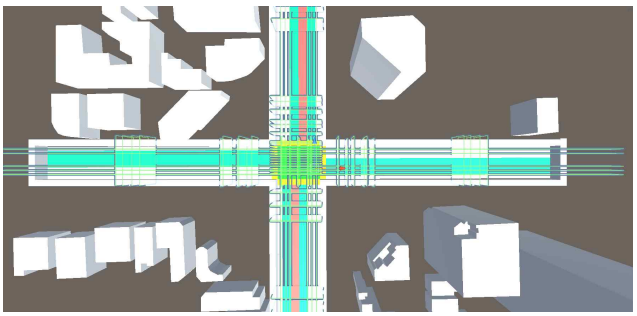


그림 3 체크포인트 구현

2.3 에이전트 설정

본 연구에서는 표 1과 같이 컴포넌트를 설정하였다.

표 1 설정한 컴포넌트

구분 및 역할	설정
Ray Perception Sensor 3D (레이어, 태그 인식)	체크포인트, 에이전트, 벽
Behavior Parameters (행동)	Discrete Branches 2 각각 Size 2, Size 4
Decision Requester	Decision Period 3
Observation	각 인식 오브젝트와의 거리

3. 보상함수 및 Epsilon

본 연구에서는 에이전트가 의도하지 않은 행동을 택할 시 음수의 보상을 얻도록 설계하였다. 따라서 에이전트가 음의 보상을 얻을 경우는 설계 의도와 다른 학습이 진행된 것으로 판단하였다.

3.1 버그 학습 분석을 통한 보상함수 설정

실제로 학습을 진행해보니 몇몇의 경우 에이전트가 최소의 음의 보상을 택하는 것을 확인하였다. 따라서 표 2와 같이 보상함수를 고려하여 조정하였다.

시행착오 내용을 모두 표 2에 기재하였고 실제 적용한 설정 값은 볼드처리 하였다.

초기 설정 상태에서도 각각의 에이전트가 교차로까지 원활하게 진입하였다. 하지만 좁은 교차로에서 각 에이전트의 충돌이 연속적으로 발생하여, 체크포인트를 지나는 가점보다 감점이 높아졌고, 이 후 에이전트가 교차로로 진입하지 않는 상황이 발생하였다.

표 2 보상함수를 고려한 설정값 변경

구분 및 역할	초기 설정	최종 설정 및 사유
타이머 감점	프레임 마다 - 0.001	**삭제** // 음의 보상 상태에서 빠르게 게임종료
벽 충돌 시 보상	-1	**-30** // 음의 보상 상태에서 빠르게 게임종료
플레이어 충돌	-5	**-0.0001** // 체크포인트를 회피하도록 유도
체크포인트 충돌	+1	

따라서 표 2와 같이 설정 값을 변화하였고 그 결과 보상 값이 수렴하는 것을 확인할 수 있었다.

3.2 Epsilon 개념 및 설정

Epsilon-greedy 알고리즘에서 Epsilon은 무작위 행동을 취한 확률을 의미한다. 이에 따라 1-Epsilon은 가장 높은 큐 함수 값을 가지는 행동을 취하도록 하는 행동을 수행할 확률을 의미한다[4].

본 연구에서는 각각의 차로에서 목적지까지 6개의 체크포인트를 생성하였다. 이에 상대적으로 체크포인트간의 거리가 길고 교차로에서의 에이전트의 탐욕적 행동을 지양하기 위해 Epsilon 값을 0.9로 설정하였다.

4. 환경 구현

자동차 모델을 에이전트로 설정, 각각의 에이전트가 위치에 따라 할당된 체크포인트를 지나 목적지까지 도달하고 다른 에이전트와 충돌하지 않도록 학습을 진행하였다.

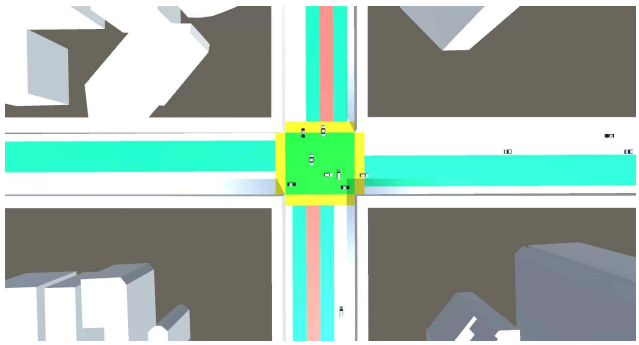


그림 4 환경 구현을 위한 학습

5. 결론 및 향후 연구

시뮬레이션을 통한 구현 및 해석의 가치는 기존 ANSYS, ABAQUS와 같은 해석코드를 통해 시간적, 경제적 측면의 효율성이 입증되었다. 드론 및 차량 자율주행의 경우에도 최적화 문제를 시뮬레이션 강화학습을 통해 해결할 수 있을 것으로 기대되는 만큼, 본 논문은 사거리와 같은 다소 복잡한 환경에도 시뮬레이션 강화학습을 적용할 수 있음을 확인하였다.

본 연구에서는 실제 도로 환경을 최대한 구현하기 위해 다른 에이전트와의 충돌을 구현하였다. 그 결과 충돌 시 Unity 물리엔진의 충돌 결과에 따라 에이전트가 택한 행동과 관계없이 에이전트의 위치 및 보상이 달라지는 노이즈 학습이 존재하는 것을 확인하였다. 에이전트간의 충돌은 감점 요소이기 때문에 학습이 진행됨에 따라 점차 충돌 횟수는 줄어들었지만 이러한 학습은 초기 무작위 탐색, 탐욕적 행동에 부정적인 요소로써 이를 개선하는 연구와 강남대로 실제 차선에 맞추어 버스 차선 및 행동 그리고 좌회전 차로를 구현하는 연구를 수행할 예정이다.

6. 참고문헌

[1] 이현호 등 3인, “유니티 기반의 항공우주분야 딥 강화학습 적용 연구”, 한국항공우주학회 학술발표최초록집, p.522-523, 2019

[2] 이성준 등 4인, “강화학습 게임에서 높은 보상 수렴 속도 개선을 위한 가상환경 기반의 DQN 모델”, 한국정보기술학회 종합학술대회 및 대학생논문경진대회, p.472-475, 2020

[3] Jack EDMONDS, “MATROIDS AND THE GREEDY ALGORITHM”, National Bureau of Standards, 1971

[4] 민규식 등 6인, “텐서플로와 유니티 ML-Agentss로 배우는 강화학습”, 위키북스, 2020