

다학제간캡스톤디자인_기말발표

2025.06.10(화)

자동차IT융합학과

20203334 김지원 / 20203336 남대현 / 20203358 이동원



1. 개요

Reward Shaping 및 Curriculum Learning 적용

Reward Shaping	Curriculum Learning
<div><div>➤ 동일한 환경에서 <u>보상함수의 구조 및 하이퍼 파라미터만 변경</u>하여 에이전트의 행동 정책을 유도</div><div>➤ <u>주행 성능 향상 및 수렴 속도 개선</u></div></div>	<div><div>➤ 학습 초기에 쉬운 환경(ex. 단순하고 장애물 개수가 적은 환경)부터 시작해 <u>점차 난도를 높여가며 학습</u>하는 전략</div><div>➤ 점진적인 일반화 능력 강화 및 <u>환경 변화에 대한 적응력 및 강건성 확보</u></div></div>

- Time-Trial에서는 트랙이 고정되고 추가적인 변수가 없으므로, Reward Shaping으로도 충분한 학습 가능
- Obstacle Avoidance나 Random Track에서는 장애물의 개수나 위치, 도로 환경 변화에 따라 학습 성능이 크게 좌우됨
- 점진적으로 환경의 난이도를 조절하는 Curriculum Learning을 적극 병행함으로써, Agent의 일반화 능력과 강건성 향상

실차 테스트 진행

- 각 Task별로 시뮬레이션에서 우수한 성능을 보인 3개의 Agent를 선별하여 실제 환경에서 테스트 진행
- 시뮬레이션에서의 성능이 실제 환경에서도 유지되는지(SimtoReal Gap)를 정량적, 정성적으로 파악하는 것이 목적
- 기존 계획했던 방법에서 비교적 간결하고 직관적으로 평가 지표를 수정

정량적 평가	정성적 평가
완주 시간, 완주 성공 횟수, 충돌 횟수	의도와 일치하는 행동, 불안정한 주행 여부, 환경 변화에 대한 강건성

2. Time Trial – 학습 전략

Time-Trial에서의 학습 전략

- 기존 보상함수의 복잡도에 비해 보상이 저조하다는 점을 고려하여, 최대한 단순하고 설계 의도가 명확하도록 구성
- Clone 기능을 활용하여, 학습의 수렴 속도를 높이기 위해 하이퍼 파라미터 조정(Reward Shaping)
- SAC의 학습 수렴 결과가 부진하였던 점을 참고하여, 단순하고 수렴성이 높은 **PPO + Continuous**로 학습 진행
Obstacle Avoidance, Random Track에도 적용

보상함수 설계

- 실제 주행에 필수적인 요소만 반영함으로써, 복잡한 조건 없이 안전성(Safety)와 효율성(Efficiency) 목표 달성
- Waypoint에 의존하지 않도록 하여 다양한 환경에서의 일반화 성능 및 강건성 확보

안전성(Safety)	효율성(Efficiency)
<div><div>➤ <u>트랙 이탈 여부</u>: 모든 바퀴가 트랙 내 있을 경우, Reward 부여</div><div>➤ <u>중심선과의 거리</u>: 트랙 중심선에 가까워질수록, 높은 Reward 부여</div><div>➤ <u>조향 각</u>: 조향각이 커질수록, 점진적으로 낮은 Reward 부여</div></div>	<div><div>➤ <u>진행률</u>: 현재 진행률만큼 Reward를 부여하여, 앞으로 나아가도록 유도</div><div>➤ <u>속도</u>: 속도에 비례하여 Reward를 부여하여, 빠르게 주행하도록 유도</div></div>

2. Time Trial – 실험 결과

테스트 결과 – (1) Time Trial Batch size 증가, Learning rate 감소

- 보상함수의 Scale을 확대하고 하이퍼 파라미터를 조정함으로써 정책이 빨리 수렴하도록 유도
- 과도한 Reward Shaping이 실차에서 과적합으로 이어질 수 있음을 고려하여, Clone 횟수를 제한적으로 설정

정량적 평가				정성적 평가	
완주 시간	lap	시뮬레이션	실차	의도와 일치하는 행동 및 불안정한 주행 여부	<div>➢ 전반적으로 보상함수의 의도와 부합하게 <u>안정적인 속도 유지</u> 및 <u>곡선 구간에서 적절한 조향</u>이 이루어졌으며, 시뮬레이션과 실차 모두 <u>일관된 주행 성능</u>을 보임</div> <div>➢ 직선 구간에서 간헐적으로 <u>지그재그 형태의 주행</u> 발생</div>
	1	13.67	15.81		
	2	14.06	15.07	시뮬레이션과 실제 주행 간 분석	<div>➢ 실차에서 시뮬레이션에 비해 속도가 다소 낮게 유지되었으며, 이는 <u>실제 환경에서의 외란(트랙 조건 및 차량의 동역학 거동)</u>의 영향을 미친 것으로 해석됨</div>
	3	13.73	14.74		
완주 성공 횟수		시뮬레이션	실차		
		3 / 3	3 / 3		

실험 결과 분석

- 곡선 구간에서 과도한 감속 없이 안정적인 속도와 조향을 유지하였음
- 직선 구간에서 지그재그 주행이 일부 발생하여, 개선 시 주행 시간이 단축할 것으로 기대됨

3. Obstacle Avoidance – 학습 전략

Obstacle Avoidance에서의 학습 전략

- Time-Trial과 마찬가지로 안전성 및 효율성 중심으로 설계하되, **장애물 회피를 우선**하도록 보상함수 설계
- 장애물의 위치나 개수에 따라 학습 성능이 크게 좌우되므로, **점진적으로 난이도를 조절하는 Curriculum Learning** 적극 병행
- Time-Trial에 비해 학습 시간을 **약 1.5배 가량 유지**하고, 수렴성이 향상하도록 **Action space 속도 범위 조절!**

속도 상한 범위가 높을수록 빠르게 주행하였음

- [1.25 2.8]로 학습 시 14.7초, [1.0 3.5]로 학습 시 14.0초 기록

시도한 방법들 – Curriculum Learning

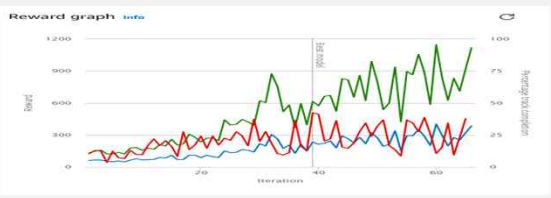
- 하나의 고정된 환경에 과적합되지 않고, **다양한 환경에서 안정적으로 동작**할 수 있도록 하는 것이 목표

장애물 개수 및 배치 방식 변화		도로 회전 방향 반전	
<div> <div>➢ 장애물 1개로 학습 후, 완주율이 일정 수준에 도달하면 점차 3~4개로 증가</div> <div>➢ 고정된 위치에 배치하여 안정적인 학습 유도 후, 추후 무작위 배치</div> </div>		<div> <div>➢ 주행 방향을 반시계에서 시계 방향으로 전환하여 코너, 장애물 조우 위치 변화</div> </div>	
한계 상황	<div> <div>➢ 장애물이 늘어남에 따라, 완주율이 저조하고 보상 경향이 불안정</div> <div>➢ 장애물 개수 2개 이상으로 무작위 배치 시, 장애물이 경로를 가로 막는 등 정상적인 주행을 방해하는 경우 발생</div> </div>	한계 상황	<div> <div>➢ 주행 방향을 유지했을 때보다, 학습 수렴까지의 시간이 증가하고 보상 경향 불안정</div> </div>

3. Obstacle Avoidance – 학습 전략

Lidar 센서 추가

- 참조한 논문(2022, 한국컴퓨터종합학술대회) 실험에 의하면, Lidar은 입력 데이터가 **많아 학습이 다소 느리지만 카메라 대비 장애물을 정확히 인식**
- 기존 단일 카메라는 장애물 인식의 정밀도가 낮고, 시각 노이즈에 취약하다고 판단하여 **이를 보완하고자 Lidar을 추가**하여 학습 시도

	Camera		Camera + Lidar	
Reward Graph		평균 Reward		평균 Reward
		200~400		800~1500
		최대 Reward		최대 Reward
		900 이상		3000 이상
수렴 경향 및 보상	➤ 약 50회차 이후 보상 증가가 둔화되고, 이후 변동 폭이 큼		➤ 약 75회차 이후 보상이 점차 증가하며, 전체적으로 상승세	

- 시뮬레이션 결과, Lidar을 추가한 모델이 수렴까지 다소 시간이 소요되었으나 **보상 추세가 안정적이며 완주율도 향상**되었음
- 실차 테스트 시, Lidar **센서의 무게로 인해 차량의 무게중심이 불안정**하여 기대한 성능을 발휘하지 못함 SimtoReal Gap

보상함수 구조

Threshold 설정

- 기존 안전성 및 효율성 보상함수의 목표와 상충됨을 고려하여, 장애물과의 **거리 기반**으로 **보상구조를 동적으로 전환**하는 방식 활용
- 장애물과의 거리가 **Threshold 이하로 진입 시 회피 행동을 유도**하도록 전환하고, 그렇지 않으면 중앙선 추종을 우선으로 함
- 기존 Threshold는 3m로 설정하였으나, 학습 시 **과도한 조향 및 경로 이탈이 빈번히 발생**하여 1.5m로 하향 조정

3. Obstacle Avoidance – 실험 결과

테스트 결과 – (2) Obstacle Avoidance

- 단일 Camera 기반으로, 장애물 개수를 1개 → 3개로 증가하거나 무작위로 배치하면서 학습
- 학습 환경의 불확실성을 고려하여, 장애물을 1개로 유지하되 무작위로 배치하면서 학습 시간을 충분히 확보하였음

정량적 평가				정성적 평가		
완주 시간	lap	시뮬레이션	실차	의도와 일치하는 행동 및 불안정한 주행 여부	➤ 시뮬레이션과의 유사한 경향을 보였으나, <u>장애물 위치와 반복 주행</u> 에 따라 불안정성이 증가함 ➤ 출발 위치에 따라 행동이 달라지는 경향이 꽤 나타났는데, 이는 <u>고정 장애물 학습의 영향</u> 으로 판단	
	1	15.42	15.12			
	2	14.67	14.64	시뮬레이션과 실제 주행 간 분석	➤ 시뮬레이터와 달리 실제 주행에서 <u>장애물 위치 차이 및 환경 요인</u> 으로 인해 정책이 일반화되지 못함 ➤ <u>곡률이 심한 구간</u> 에서 장애물을 조우할 때 회피 실패하였으며, 위의 이유로 2번째 장애물과 <u>충돌 발생</u>	
	3	13.59	14.00			
완주 성공 횟수 [충돌 횟수]		시뮬레이션	실차			
		3 / 3 [0]	3 / 3 [2]			

실험 결과 분석

속도 유지 / 장애물 회피 / 조향 최소화

- 트랙 안쪽에서 출발 시, 1번째와 3번째 장애물을 안정적으로 회피하며 완주 성공 → 보상 함수의 의도가 실제 주행에서 잘 반영됨
- 트랙 중앙에서 출발 시, 2번째 장애물의 위치가 시뮬레이션과 달라 충돌이 발생하였으며 2번째 lap에서도 동일한 현상이 반복됨
 - 특정 장애물 위치에 과적합되거나 반복 주행에 대한 회피 전략이 미흡함을 시사


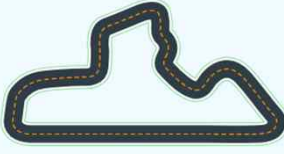
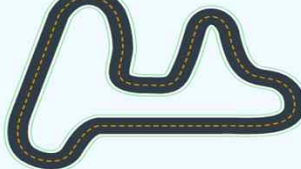
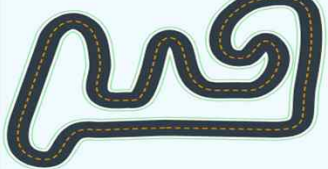
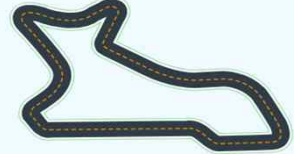
4. Random Track – 학습 전략

Random Track에서의 학습 전략

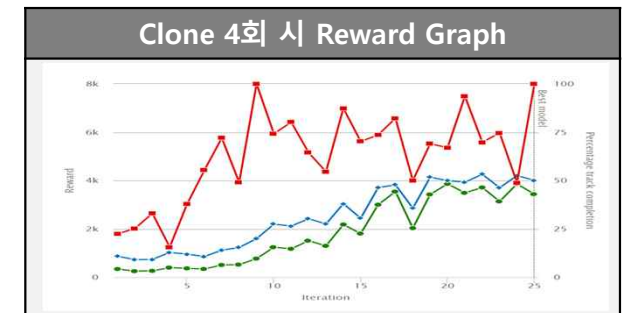
- Time-Trial에서와 마찬가지로, 트랙에 종속적이지 않은 간결하고 명확한 보상구조를 설계하여 일반화 성능 확보
- 곡률이 작고 간단한 트랙으로부터 곡률이 크고 난도가 높은 트랙으로 확장하는 Curriculum Learning 적용

Curriculum Learning 학습

- 각 Clone(1시간 30분)마다 트랙 구조를 점진적으로 변화하였으며, 가능한 트랙의 구조가 중복되지 않도록 선별하였음

Clone 횟수	초기 모델	1회	2회	3회	4회
트랙 형태					

- 시뮬레이션 시, Clone을 4회 진행한 Agent의 보상 수렴 정도나 완주율 측면에서 가장 성능이 우수함
- 실차 테스트에서는 오히려 Clone을 **1회 진행한 Agent가 더 좋은 성능**을 보임
 - 실차 환경이 시뮬레이션에 비해 이상적이지 않음(ex. 센서의 노이즈, 트랙의 마찰력 등) **SimtoReal Gap**
 - Curriculum Learning이 일정 수준까지는 효과가 명확하나, 그 이상에서는 추가적인 성능 향상 미미



4. Random Track – 실험 결과

테스트 결과 – (3) Random Track

- Time-Trial에서의 보상함수를 그대로 유지하면서, 하이퍼 파라미터를 조정하고 점차 곡률이 크고 복잡한 트랙으로 확장
- Curriculum Learning 수준과 SimtoReal Gap의 영향을 고려하여, 실차 테스트에서는 Clone 1회 Agent를 채택

정량적 평가				정성적 평가	
완주 시간	lap	시뮬레이션	실차	의도와 일치하는 행동 및 불안정한 주행 여부	> 세 Task 중 차선 중앙을 가장 <u>안정적으로 추종</u> > 다만, 속도가 높을 경우 곡률이 큰 구간에서 반복적으로 <u>차선 이탈이 발생함</u>
	1	11.15	9.84		
	2	12.47	9.37	시뮬레이션과 실제 주행 간 분석	> 실차 주행에서 더 빠르고 안정적인 주행 성능을 보임 > 시뮬레이션과 실차 간의 환경 차이로 인해 <u>성능 예측이 어려움을 보여주는 사례</u> 로 해석됨
	3	11.94	9.04		
완주 성공 횟수		시뮬레이션	실차		
		3 / 3	3 / 3		

실험 결과 분석

- 사전에 학습하지 않은 트랙에서도 안정적으로 주행한 것으로 보아, 설계한 Agent가 의도한 대로 높은 강건성을 확보하였음을 확인
- 간헐적으로 속도가 높을 경우, 곡률이 큰 구간에서 트랙 이탈이 발생 고속에서 조향 제어의 한계
 → 보상함수 설계 또는 정책 안정화 측면에서 개선 필요성 시사

5. SimtoReal Gap 분석 및 고찰

SimtoReal Gap

- 정량적 및 정성적 지표에서 상당한 차이가 존재하였으며, 설계한 보상함수와 학습 전략이 실차 주행에 충분히 반영되지 못함
 - 1) 실제 트랙의 마찰력과 센서 노이즈 등은 시뮬레이터처럼 이상적이지 않음
 - 2) 실차는 무게 중심, 관성 등(차량 동역학 오차)으로 인해 시뮬레이터처럼 정확히 차선 중앙을 추종하지 못함
- 시뮬레이션에서 완주하지 못한 Agent가 실차에서 완주한 경우가 있었으나, 이는 환경 요소와 우연히 맞물린 결과로 일반화 가능성 낮음
- DeepRacer 플랫폼은 설계와 사전 검증에 유용하나, 실차 적용을 위한 신뢰성 확보 측면에서는 한계가 존재함

개선 방안

- DeepRacer 플랫폼으로 진행하기에는 한계가 있으므로, ROS나 Gazebo 등을 활용할 수 있음

동역학 및 센서 정합성 향상	<ul style="list-style-type: none">차량 질량, 무게 중심, 마찰 계수 등 <u>동역학 요소</u>와 <u>센서 노이즈</u>를 반영하여 현실성 향상<u>Gazebo</u>를 활용하여 실차와의 정합성 확보
도메인 랜덤화	<ul style="list-style-type: none">트랙 내 조명, 텍스처, 날씨 등 <u>시각적 요소</u>를 <u>랜덤화</u> 하여 시뮬레이션 다양성 확보<u>ROS</u> 및 <u>Gazebo</u> 연동으로 환경 조건을 유연하게 설정
강화학습 외 경로 추종 알고리즘 활용	<ul style="list-style-type: none">주행 오차(횡 방향 및 각도)를 수집하여 <u>경로 추종 제어(Pure Pursuit, Stanley)</u> 적용<u>ROS</u>를 통한 성능 평가 및 제어