2025학년도 1학기 SW 캡스톤디자인 경진대회

강화학습기반의 자율이동로봇의 최적경로 알고리즘

팀 명 패스파인더

지도교수 박영진

팀 원 (it지능정보공학과, 정우현),(it정보공학과, 김준혁),
(컴퓨터인공지능학부, 김경만), (컴퓨터인공지능학부, 이성주)

산업체 오토메스텔스타

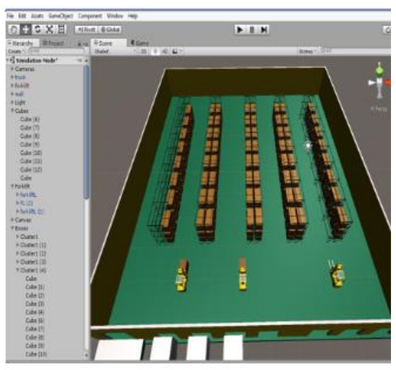
개발 동기 및 목적

1. 개발 동기 및 목적

배경: 물류 자동화와 스마트 팩토리 환경에서 **작업량 증가**와 **배터리 소모 최소화**가 중요한 과제로 부상.

문제점: 기존 ROS 기반 자율이동로봇(AMR)은 경로를 일일이 지정해야 하며, **교착 상태**나 **이동 장애물(사람 등)**에 취약.

목적: 강화학습 기반 경로 최적화 알고리즘을 통해 AMR이 **스스로 최적 경로를 학습하고**, 교착 및 장애물을 **회피**할 수 있게 하기 위함.



주요 기술

2. 주요 기술

1) 강화학습 프레임워크 – Unity ML-Agents

Unity ML-Agents Toolkit을 사용하여 시뮬레이션 기반의 디지털 트윈 환경을 구현.

실제 로봇을 학습시키지 않고도 다양한 시나리오에서 자율이동로봇(AMR)을 훈련시킬 수 있음.

3D 물리 시뮬레이션을 통해 실제 환경에 근접한 학습 가능.

2) 강화학습 알고리즘 – PPO (Proximal Policy Optimization)

OpenAI에서 제안한 최신 정책 기반 강화학습 알고리즘.

기존의 DQN, A3C보다 안정적인 수렴 특성과 효율성을 가짐.

정책 변화 폭에 제한(Clip)을 걸어 급격한 정책 변경을 방지, 학습 안정성 확보.

학습 과정 중 과거의 정책을 기준으로 보상 함수를 조절함으로써 부정적인 보상 누적을 피함.

3) 디지털 트윈(Digital Twin)

실제 AMR 환경을 Unity에서 가상으로 정밀하게 구현하여 현실과 같은 조건에서 학습.

현실 환경에서 발생 가능한 변수(장애물, 교착 상황 등)를 사전에 고려한 훈련이 가능.

4) 센서 기술 적용

Grid Sensor: 로봇을 중심으로 주변을 격자 형태로 감지하여 위치 정보를 파악.

Ray Sensor: 전방 및 주변에 광선을 쏘아 벽, 장애물, 목표물의 방향과 거리를 인식.

다양한 센서를 결합하여 에이전트가 보다 풍부한 상태 정보를 받아 판단.

5) 커리큘럼 러닝(Curriculum Learning)

학습 초기에는 쉬운 환경(장애물 적음)에서 시작하고, 점진적으로 난이도를 증가시킴.

난이도 조절 요소: **벽의 수**, **AMR 수**, **목표물의 위치 난이도**

이를 통해 에이전트가 학습 중간에 좌절하지 않고 효과적으로 정책을 학습하도록 유도.

6) 멀티에이전트 학습 전략 – 독립학습 방식

에이전트 간 정책을 공유하지 않고 각자 독립적으로 학습.

상대방을 환경의 일부로 인식하여 교착 상태 회피 및 협력 아닌 경쟁 상황에 유리.

상호 간섭을 최소화하고 현실 로봇 적용 가능성을 높임

개발 내용

개발 내용

보상 설정:

목적지 도달: 보상
벽 충돌: 패널티
시간 경과: 패널티
이동 시: 보상 (정지 방지)
목표에 가까워질수록: 보상

강화학습 구성 요소:

상태(state): 위치, 장애물, 속도
행동(action): 이동 방향
보상(reward): 행동 피드백
에이전트(agent): AMR 본체

학습 전략 개선:

커리큘럼 러닝: 점진적으로 맵 난이도 상승
(벽/AMR 수 조절)
센서 감지 범위 확장: 전방향 인식 → 부딪힘 문제 감소
에이전트 조작 개선: 미끄러짐 감소로 학습 안정화

학습 방식:

독립 학습 방식 선택 (경쟁 학습보다 안정적

결과 및 분석

4. 결과 및 분석

에피소드 2,500,000번 학습 결과:

목표 지점 도달 성공률 향상
다양한 상황에서 **최적 경로를 찾는 능력 향상**

한계:

교착 상태나 긴 벽 앞에서 여전히 어려움
학습량만 늘린다고 개선되진 않음 →
보상 정책과 학습 환경의 정교화 필요

해결 방향:

다양한 맵에서 **의도적으로 교착 상황 생성**
→ 학습에 반영
보상 정책 정제 및 **에피소드 다양화** 필요