

AI GAMES

2023622076 정현수

2023864019 김현진



강화학습 기반 미로 탈출 대전게임

프로젝트 개요:

- AI1 : Q-Learning 학습 에이전트
- AI2 : 규칙 기반 에이전트
- 목표 : 조건 달성 후 골 도달

사용된 기술 스택

- Python + Pygame
- Q-Learning 알고리즘
- Supabase 데이터베이스

게임 규칙

승리 조건

- 코인 3개 수집
 - 아이템 1회 적중
 - 체크포인트 도달
- 🏆 골 연락 & 도달

맵 구성

- 벽(1): 이동 불가
- 포탈(2): 순간이동 (2개 페어)
- 게이트(3): 골 연락 시 차단
- 체크포인트(4): 필수 경유지
- 골(5): 최종 목적지

아이템 시스템

- 코인: 맵에 3개 랜덤 스폰
- 트랩: 설치형, 상대 5턴 스텐
- 마취총: 인접 자동 발사, 5턴 스텐
- 획득 후 재스폰

게임 제한

- 턴 제한: 240턴
- 제한 초과 시: 무승부
- 골 도달 시: 즉시 승리

게임 인터페이스

난이도 설정

EASY	0.7	AI10 70% 랜덤 선택
NORMAL	0.3	AI10 30% 랜덤 선택
HARD	0.05	AI10 5% 랜덤 선택

1. 🤖 AI vs AI

- AI1 (Q-Learning) vs AI2 (규칙)
- 자동 플레이 관전 모드
- 빠른 학습 검증 가능
- 통계 실시간 확인

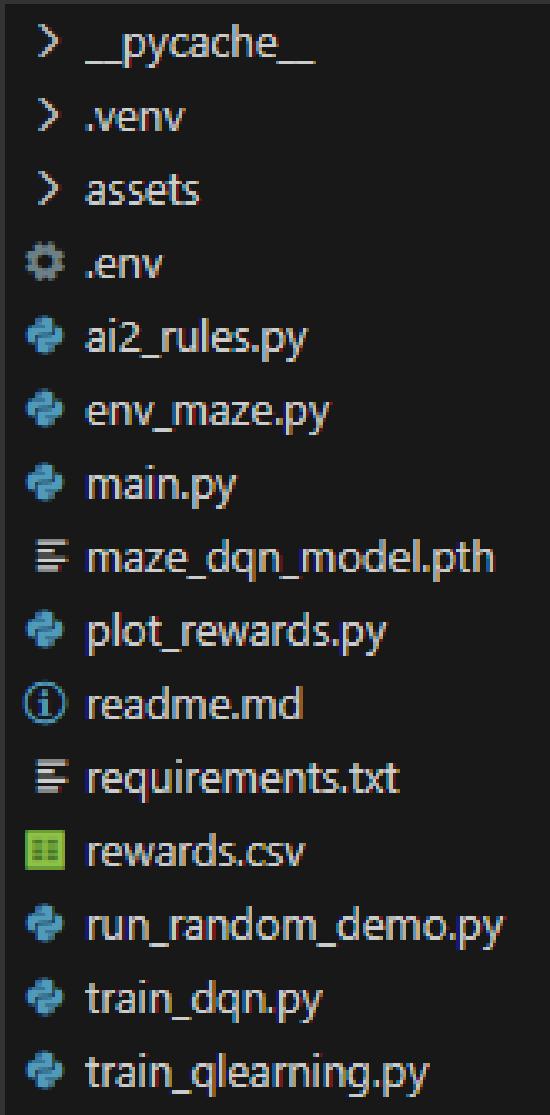
2. 🧑 AI vs Player

- 플레이어가 AI2 조작
- WASD/화살표: 이동
- E: 트랩 설치
- 학습된 AI와 대결

주요 컨트롤

- P : 일시정지 / 재개
- ESC : 타이틀로 돌아가기
- F : 빠른 모드 (AI vs AI)
- M : 30 라운드 스킵 (AI vs AI)
- N : 다음 게임 (AI vs AI)

주요 시스템 구조



env_mazes 게임 환경 (미로, 물리, 렌더링)

ai2_rules AI2 규칙 기반 로직 (BFS)

train_qlearning AI1 Q-Learning 학습

main 게임 실행 (Pygame GUI)

기술적 세부사항

env_maze.py 핵심 기능

- 그리드 시스템 : 15×11 (720×528 픽셀)
- 셀 크기 : 48×48 픽셀
- 턴 제한 : 240턴 (약 24초@10FPS)
- 스텐 지속시간 : 5턴

게이트 메커니즘

- 평소 : 통행 가능 (파란색)
- goal_activate = True: 벽처럼 차단 (어두운 파란색)
- 한쪽이 언락하면 양쪽 모두 영향

포탈 시스템

- 2개의 포탈이 페어로 연결
- (1, 0) , (13, 10)
- 즉시 순간이동 (딜레이 X)

아이템 재스폰 로직

- 코인 : 획득 즉시 1개 재스폰 (항상 3개 유지)
- 트랩 픽업 : 누군가 획득 시 즉시 재스폰
- 마취총 픽업 : 누군가 획득 시 즉시 리스폰

AI2 : 규칙 기반 에이전트

① 행동 우선 순위

1. 코인 < 3개 → 가장 가까운 코인으로 이동
2. 코인 충분 & 아이템 미적중:
 - 아이템 없음 → 트랩/마취총 획득
 - 아이템 있음 → 상대에게 접근 & 공격
3. 체크포인트 미도달 → 체크포인트로 이동
4. 골 언락 완료 → 골로 이동

② 경로 탐색:

- BFS (Breadth-First Search) 알고리즘 사용
- 최단 경로 자동 계산
- 상대 트랩 회피 로직
- 벽과 장애물 고려

③ 난이도 조절:

- 기본 랜덤 : 7%
- 쓸데없는 이동 : 18%
- 총 약 25% 비최적 행동
- AI1이 학습 가능한 수준

④ 특징

- 일관성 있는 전략적 플레이
- 목표 지향적 행동
- AI1의 학습 대상으로 적합

Q-Learning 학습

STATE

- 위치 (x, y)
- 코인 개수: 0 ~ 3
- 아이템 적중 여부 : T/F
- 체크포인트 도달 : T/F
- 골 언락 여부 : T/F

학습 파라미터

- 에피소드 : 50,000회
- 학습률 : 0.1
- 할인율 : 0.95
- Epsilon : 1.0 (시작) → 0.01 (최소)
- Decay : 0.9995

보상 설계

- 코인 획득 : +80
- 아이템 적중 : +80
- 체크포인트 : +100
- 골 언락 : +100
- 승리 : + 300
- 패배 : - 100

학습 프로세스

환경 초기화 → 상태 관찰 → 행동 선택 (ϵ -greedy) → 환경 실행 → Q값 업데이트

Supabase 연동

- 2,000 에피소드마다 중간 저장
- 학습 중단 시에도 데이터 보존
- Q값 0인 데이터는 제외 (용량 절약)
- 배치 업로드 (1,000개씩)

데이터 형식

- state : 문자열 (튜플 표현)
- action : 정수 (0~3)
- q_value : 실수

```
=====
```

[START] 최적화된 Q-Learning 학습 시작

```
=====
```

에피소드 : 50000

Max Steps: 300 (게임 제한: 240)

Epsilon: 1.0 → 0.01 (decay: 0.9995)

중간 저장 : 2000 에피소드마다

```
=====
```

Ep	20	Reward: -105.10 (avg: -52.04)	Win Rate: 0.0%	Steps: 103 (avg: 104.9)	ϵ : 0.9900	Q-size: 444	Time: 0s
Ep	40	Reward: 54.65 (avg: -36.85)	Win Rate: 0.0%	Steps: 110 (avg: 105.3)	ϵ : 0.9802	Q-size: 932	Time: 0s
Ep	60	Reward: 230.56 (avg: -29.11)	Win Rate: 0.0%	Steps: 87 (avg: 101.5)	ϵ : 0.9704	Q-size: 1204	Time: 0s
Ep	80	Reward: 134.90 (avg: -25.47)	Win Rate: 0.0%	Steps: 106 (avg: 107.4)	ϵ : 0.9608	Q-size: 1540	Time: 0s
Ep	100	Reward: 54.90 (avg: -24.32)	Win Rate: 0.0%	Steps: 105 (avg: 109.4)	ϵ : 0.9512	Q-size: 1700	Time: 0s
Ep	120	Reward: -25.90 (avg: -22.62)	Win Rate: 0.0%	Steps: 120 (avg: 110.7)	ϵ : 0.9418	Q-size: 1844	Time: 1s
Ep	140	Reward: 134.20 (avg: -24.92)	Win Rate: 0.0%	Steps: 120 (avg: 110.7)	ϵ : 0.9324	Q-size: 1904	Time: 1s
Ep	160	Reward: -27.80 (avg: -32.81)	Win Rate: 0.0%	Steps: 158 (avg: 112.7)	ϵ : 0.9231	Q-size: 1952	Time: 1s
Ep	180	Reward: -24.05 (avg: -24.83)	Win Rate: 0.0%	Steps: 83 (avg: 109.7)	ϵ : 0.9139	Q-size: 2088	Time: 1s
Ep	200	Reward: -104.55 (avg: -22.37)	Win Rate: 0.0%	Steps: 92 (avg: 105.5)	ϵ : 0.9048	Q-size: 2164	Time: 1s
Ep	220	Reward: -104.65 (avg: -17.75)	Win Rate: 0.0%	Steps: 94 (avg: 105.3)	ϵ : 0.8958	Q-size: 2184	Time: 1s
Ep	240	Reward: -25.95 (avg: -15.98)	Win Rate: 0.0%	Steps: 121 (avg: 105.0)	ϵ : 0.8869	Q-size: 2388	Time: 1s
Ep	260	Reward: 53.45 (avg: -10.09)	Win Rate: 0.0%	Steps: 134 (avg: 105.3)	ϵ : 0.8781	Q-size: 2408	Time: 1s
Ep	280	Reward: -102.85 (avg: -20.66)	Win Rate: 0.0%	Steps: 58 (avg: 104.8)	ϵ : 0.8693	Q-size: 2484	Time: 1s
Ep	300	Reward: -25.75 (avg: -23.52)	Win Rate: 0.0%	Steps: 117 (avg: 108.9)	ϵ : 0.8607	Q-size: 2512	Time: 1s
Ep	320	Reward: -105.50 (avg: -26.74)	Win Rate: 0.0%	Steps: 111 (avg: 107.5)	ϵ : 0.8521	Q-size: 2616	Time: 1s
Ep	340	Reward: 335.49 (avg: -21.68)	Win Rate: 0.0%	Steps: 240 (avg: 109.0)	ϵ : 0.8436	Q-size: 2720	Time: 1s
Ep	360	Reward: -106.10 (avg: -27.34)	Win Rate: 0.0%	Steps: 123 (avg: 109.2)	ϵ : 0.8352	Q-size: 2748	Time: 2s
Ep	380	Reward: -24.25 (avg: -30.58)	Win Rate: 0.0%	Steps: 87 (avg: 107.7)	ϵ : 0.8269	Q-size: 2776	Time: 2s
Ep	400	Reward: 55.25 (avg: -27.72)	Win Rate: 0.0%	Steps: 98 (avg: 104.3)	ϵ : 0.8187	Q-size: 2864	Time: 2s
Ep	420	Reward: -104.70 (avg: -23.49)	Win Rate: 0.0%	Steps: 95 (avg: 104.9)	ϵ : 0.8105	Q-size: 2880	Time: 2s
Ep	440	Reward: -105.95 (avg: -33.57)	Win Rate: 0.0%	Steps: 120 (avg: 104.5)	ϵ : 0.8025	Q-size: 2880	Time: 2s
Ep	460	Reward: -106.35 (avg: -36.77)	Win Rate: 0.0%	Steps: 128 (avg: 105.6)	ϵ : 0.7945	Q-size: 2880	Time: 2s
Ep	480	Reward: -103.55 (avg: -32.53)	Win Rate: 0.0%	Steps: 72 (avg: 104.9)	ϵ : 0.7866	Q-size: 2892	Time: 2s
Ep	500	Reward: 170.71 (avg: -35.24)	Win Rate: 0.0%	Steps: 97 (avg: 106.8)	ϵ : 0.7788	Q-size: 2948	Time: 2s
Ep	520	Reward: -103.55 (avg: -27.42)	Win Rate: 0.0%	Steps: 72 (avg: 107.4)	ϵ : 0.7710	Q-size: 3012	Time: 2s
Ep	540	Reward: 180.57 (avg: -22.17)	Win Rate: 0.0%	Steps: 129 (avg: 108.5)	ϵ : 0.7633	Q-size: 3016	Time: 2s
Ep	560	Reward: -104.10 (avg: -16.77)	Win Rate: 0.0%	Steps: 83 (avg: 106.3)	ϵ : 0.7557	Q-size: 3016	Time: 2s
Ep	580	Reward: -108.65 (avg: -16.44)	Win Rate: 0.0%	Steps: 174 (avg: 110.2)	ϵ : 0.7482	Q-size: 3016	Time: 2s
Ep	600	Reward: 55.90 (avg: -17.04)	Win Rate: 0.0%	Steps: 85 (avg: 108.8)	ϵ : 0.7408	Q-size: 3016	Time: 3s
Ep	620	Reward: -27.15 (avg: -25.78)	Win Rate: 0.0%	Steps: 145 (avg: 107.3)	ϵ : 0.7334	Q-size: 3016	Time: 3s
Ep	640	Reward: -107.90 (avg: -26.11)	Win Rate: 0.0%	Steps: 159 (avg: 107.8)	ϵ : 0.7261	Q-size: 3028	Time: 3s
Ep	660	Reward: 134.74 (avg: -23.30)	Win Rate: 0.0%	Steps: 118 (avg: 105.3)	ϵ : 0.7189	Q-size: 3028	Time: 3s
Ep	680	Reward: -104.85 (avg: -36.19)	Win Rate: 0.0%	Steps: 98 (avg: 103.5)	ϵ : 0.7117	Q-size: 3028	Time: 3s
Ep	700	Reward: -24.30 (avg: -33.56)	Win Rate: 0.0%	Steps: 88 (avg: 103.5)	ϵ : 0.7046	Q-size: 3124	Time: 3s
Ep	720	Reward: -104.50 (avg: -35.03)	Win Rate: 0.0%	Steps: 91 (avg: 102.5)	ϵ : 0.6976	Q-size: 3152	Time: 3s
Ep	740	Reward: -24.35 (avg: -37.99)	Win Rate: 0.0%	Steps: 89 (avg: 98.3)	ϵ : 0.6907	Q-size: 3152	Time: 3s
Ep	760	Reward: -23.35 (avg: -38.17)	Win Rate: 0.0%	Steps: 69 (avg: 101.8)	ϵ : 0.6838	Q-size: 3152	Time: 3s
Ep	780	Reward: 54.70 (avg: -21.47)	Win Rate: 0.0%	Steps: 109 (avg: 103.5)	ϵ : 0.6770	Q-size: 3152	Time: 3s
Ep	800	Reward: -27.10 (avg: -33.98)	Win Rate: 0.0%	Steps: 144 (avg: 104.8)	ϵ : 0.6703	Q-size: 3152	Time: 3s
Ep	820	Reward: 135.40 (avg: -30.22)	Win Rate: 0.0%	Steps: 96 (avg: 106.0)	ϵ : 0.6636	Q-size: 3160	Time: 3s
Ep	840	Reward: 55.20 (avg: -26.38)	Win Rate: 0.0%	Steps: 99 (avg: 107.7)	ϵ : 0.6570	Q-size: 3180	Time: 3s

Ep 49820	Reward: 268.44 (avg: 285.34)	Win Rate: 13.0%	Steps: 102 (avg: 152.6)	ϵ : 0.0100	Q-size: 8500	Time: 262s
Ep 49840	Reward: 153.95 (avg: 291.04)	Win Rate: 16.0%	Steps: 125 (avg: 141.4)	ϵ : 0.0100	Q-size: 8500	Time: 262s
Ep 49860	Reward: 317.84 (avg: 283.47)	Win Rate: 15.0%	Steps: 89 (avg: 140.4)	ϵ : 0.0100	Q-size: 8500	Time: 262s
Ep 49880	Reward: 157.30 (avg: 257.44)	Win Rate: 12.0%	Steps: 58 (avg: 129.7)	ϵ : 0.0100	Q-size: 8500	Time: 262s
Ep 49900	Reward: 76.80 (avg: 265.78)	Win Rate: 14.0%	Steps: 67 (avg: 121.9)	ϵ : 0.0100	Q-size: 8500	Time: 262s
Ep 49920	Reward: 149.45 (avg: 265.09)	Win Rate: 15.0%	Steps: 215 (avg: 119.2)	ϵ : 0.0100	Q-size: 8500	Time: 262s
Ep 49940	Reward: 336.00 (avg: 255.56)	Win Rate: 11.0%	Steps: 84 (avg: 127.6)	ϵ : 0.0100	Q-size: 8500	Time: 262s
Ep 49960	Reward: 76.25 (avg: 271.45)	Win Rate: 13.0%	Steps: 78 (avg: 134.2)	ϵ : 0.0100	Q-size: 8500	Time: 262s
Ep 49980	Reward: 328.20 (avg: 275.10)	Win Rate: 13.0%	Steps: 240 (avg: 141.6)	ϵ : 0.0100	Q-size: 8500	Time: 262s
Ep 50000	Reward: 152.85 (avg: 266.38)	Win Rate: 11.0%	Steps: 147 (avg: 145.5)	ϵ : 0.0100	Q-size: 8500	Time: 262s

=====

[CHECKPOINT] 50000 에피소드 완료 - 중간 저장 중...

[INFO] Supabase 테이블 'q_table_maze_v4'에 데이터를 저장하는 중...

> 1000 / 7134 개 저장 완료...
> 2000 / 7134 개 저장 완료...
> 3000 / 7134 개 저장 완료...
> 4000 / 7134 개 저장 완료...
> 5000 / 7134 개 저장 완료...
> 6000 / 7134 개 저장 완료...
> 7000 / 7134 개 저장 완료...
> 7134 / 7134 개 저장 완료...

[SUCCESS] 모든 데이터가 성공적으로 저장되었습니다! (총 7134개)

=====

[FINAL SAVE] 최종 Q-table 저장 중...

[INFO] Supabase 테이블 'q_table_maze_v4'에 데이터를 저장하는 중...

> 1000 / 7134 개 저장 완료...
> 2000 / 7134 개 저장 완료...
> 3000 / 7134 개 저장 완료...
> 4000 / 7134 개 저장 완료...
> 5000 / 7134 개 저장 완료...
> 6000 / 7134 개 저장 완료...
> 7000 / 7134 개 저장 완료...
> 7134 / 7134 개 저장 완료...

[SUCCESS] 모든 데이터가 성공적으로 저장되었습니다! (총 7134개)

=====

[학습 완료 통계]

총 에피소드 : 50000
총 학습 시간: 264.0초 (4.4분)
최종 승률: 11.0% (최근 100개임)
평균 보상: 266.38 (최근 100개임)
Q-table 크기: 8500 상태-행동 쌍

=====

AI Q-Learning 의사 결정

현재 상태 확인

Q-table에서 4가지 행동의 Q값 조회

ϵ -greedy 정책 적용

- ϵ 확률 : 랜덤 행동 (탐험)
- $1-\epsilon$ 확률 : 최대 Q값 행동 (활용)

행동 실행

보상 받기 + 다음 상태 관찰

Q값 업데이트

감사합니다