**인공지능 과제2 : Q-Learning**

2018062733 윤동빈

1. 함수 설명

5\*5 배열에서 행과 열을 각각 x, y로 나타낸 좌표로 표현했을 때,

텍스트, 장치, 측정기, 게이지이(가) 표시된 사진

자동 생성된 설명

OOB(x, y) : 주어진 5\*5 맵을 벗어나는 경우를 체크하는 함수

텍스트이(가) 표시된 사진

자동 생성된 설명

state(x, y) : (x, y)좌표를 맵에서의 위치(state)로 변환하는 함수

2. 코드 설명

텍스트이(가) 표시된 사진

자동 생성된 설명

dx, dy : 현재 위치 (x, y)에서 방향 a(0~3)가 주어졌을 때 다음 위치를 계산하기 위한 변수  
q : state s, 방향 a가 주어졌을 때 Q(s, a)를 의미  
reward : (x, y)에서의 reward  
bomb : (x, y)에 폭탄이 있는지 여부  
x, y : 현재 좌표  
start\_x, start\_y : 시작 지점 좌표

텍스트이(가) 표시된 사진

자동 생성된 설명5\*5 맵 정보가 담긴 input.txt를 가져와  
reward, bomb에 반영하는 과정

Q-Learning 구현 코드

텍스트이(가) 표시된 사진

자동 생성된 설명현재 위치를 시작 지점으로 초기화  
  
loop를 50만번 진행(cnt) : 수렴 조건

현재 위치에 폭탄이 있다면 시작 위치로 이동  
  
다음 방향은 랜덤으로 설정

정해진 방향(a) 정보를 통해 다음 위치 설정

맵을 벗어나지 않을 때까지  
방향 재설정

다음 state s’에서 모든 방향에 대해  
Q(s’, a’) 최댓값 구하기

Q(s, a) 갱신

다음 state로 이동

텍스트이(가) 표시된 사진

자동 생성된 설명output.txt 작성하는 과정

시작 위치의 Q(s, a) 최댓값 구하기

현재 위치를 시작 위치로 설정

현재 state 기록

Goal state 도착 시  
시작 위치의 Q(s, a) 최댓값 기록 및 종료

현재 위치에서 가장 큰 Q(s, a)를 갖는  
방향 a를 찾아 다음 위치(nx, ny)로 진행

3. 실험 결과 및 분석

reward = 1일 때 output.txt  


reward = 10일 때 output.txt  


reward = 20일 때 output.txt  


reward가 1부터 20까지 계속 증가했음에도 시작 지점부터 도착 지점까지의 **경로 변화는 없었음**  
하지만 시작 지점에서 도착 지점까지의 이동 가능 경로 주위에 bonus point가 위치해 있어 시작 지점의 최대 Q(s, a)는 증가  
그럼에도 경로 변화가 없었던 이유는 **상대적으로 큰 도착 지점의 reward 영향력**이 컸기 때문