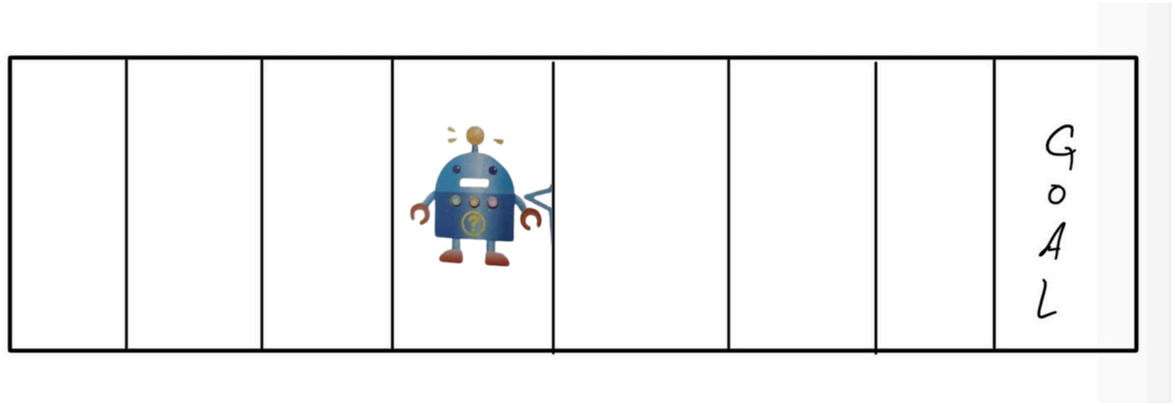
강화학습과 최적화

대전대신고등학교 30602 강은석

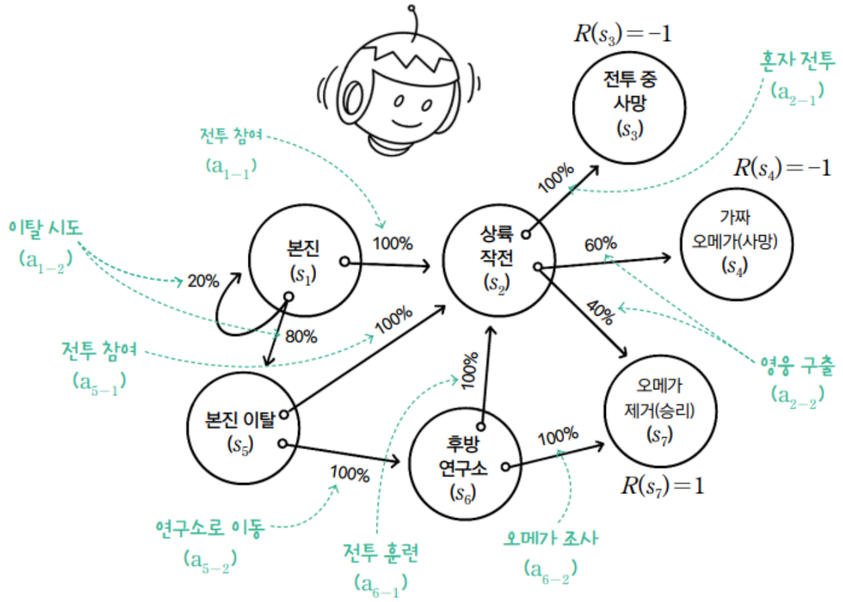
목차

강화학습이란?

머신러닝의 한 분야로, 의사 결정에 주로 사용되는 방법



Markov Decision Process



$$p(a_1|s_0a_1s_1)$$

$$p(a_1|s_1)$$

강화학습의 종류

Value-based

가치 기반

Policy-based

정책 기반

Value-based

목적 함수로 상태 가치 함수, 행동 가치 함수를 설정. 보통 행동 가치 함수를 최대화하는 정책이 최적 정책이다.

상태 가치 함수 $$V(s) = E[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t) | s_0 = s]$$

Value-based

목적 함수로 상태 가치 함수, 행동 가치 함수를 설정. 보통 행동 가치 함수를 최대화하는 정책이 최적 정책이다.

행동 가치 함수 $$Q(s,a) = E[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t) | s_0 = s, a_0 = a]$$

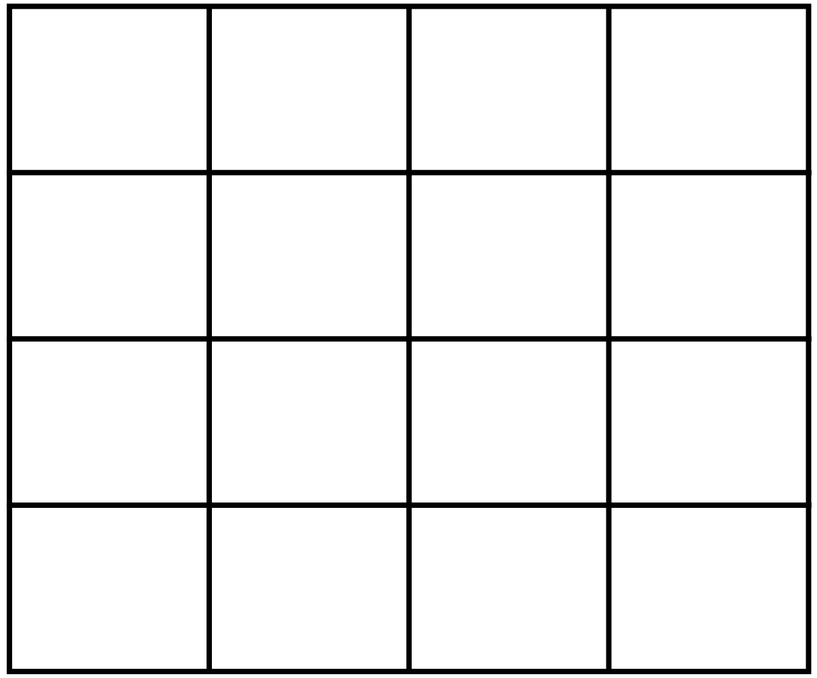
Bellman-equation

상태 가치 함수, 행동 가치 함수를 재귀적으로 정의

상태 가치 함수 $V(s) = E[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t) | s_0 = s]$

Q-learning

대표적인 가치 기반 강화학습으로, Q(행동 가치 함수)를 최대화하는 방향으로 학습한다.

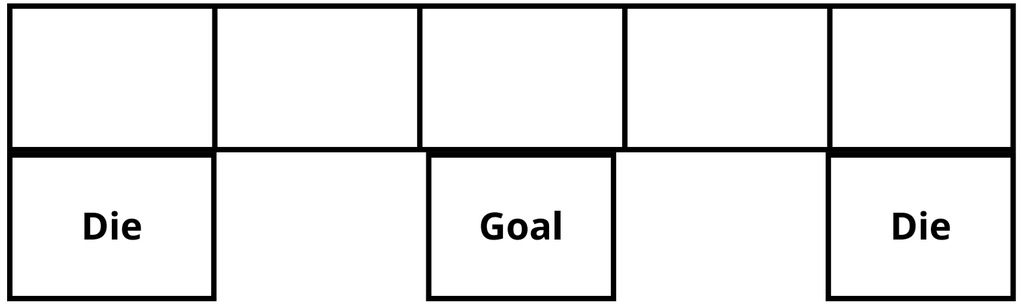


Policy-based

가치 기반의 문제

1) 연속적인 문제에 적용하기 힘듦

2) 상황에 대한 행동이 한 개로 결정



Policy-based

정책 기반의 특징

1) 연속적인 문제에 적용 가능
2) 상황에 대한 행동이 평균과 분산으로 결정

Policy-based

보상 자체를 목적함수로 두어 사용

$$J \equiv E[G_0] = \int_{s_0, a_0, \cdots} G_0 P(s_0, a_0, \cdots) d s_0, a_0, \cdots$$

Optimization

Optuna	정책기반
C : 4.77	C : 20.01
γ : 0.21	γ : 0.01
소요 시간 : 3초	소요 시간 : 18분 21초
스코어 : 0.9867	스코어 : 0.9733

느낀 점

강화학습의 개념과 종류에 대해 알게 되었고, 단순한 최적화 문제에서는 강화학습보 다 optuna와 같은 비교적 간단한 방법이 더욱 효율적일 수 있다는 사실을 알게 되었다.

Thank you.