# 강화학습 분류체계

- 강화학습은 여러 분류 기준이 존재한다.

## MDP를 알고있는지 여부

### MDP를 알때

- 환경의 전이 확률과 보상 함수를 알고있다
- 정책반복(Policy Iteration)과 가치반복(Value Iteration)과같은 동적 프로그래밍 방법을 사용할수 있다.

### MDP를 모를때

- 환경의 전이확률과 보상 함수를 모르고있다.
- 샘플 기반 알고리즘을 사용한다. MC, TD

## 가치기반, 정책 기반

### 가치기반

- 가치 함수를 최적화 하는 방법
- 최적의 가치 함수를 찾은 후 이를 기반으로 최적의 정책을 유도한다.
- Q-learning

### 정책기반

- 정책 자체를 파라미터화 시켜 최적화 하는 방법
- 행동 공간이 연속 적일때 유용하다.
- 정책 그래디언트

## 싱글, 멀티 에이전트

### 싱글

- 환경과 상호작용 하는 에이전트는 하나이다.

### 멀티

- 여러에이전트가 상호 작용 하며 학습한다.
- 협력 및 경쟁 관계가 복잡한 문제에 사용된다.

## 연속, 이산행동 공간

### 연속

- 행동이 연속적인 값으로 표현되는 경우
- 정책 기반 방법, 엑터-크리틱 방법

### 이산

- 행동이 이산적인 값으로 표현되는 경우
- 가치 기반 방법

## 모델기반, 모델 프리

### 모델기반

- 환경모델을 사용하거나 학습하며 예측과 계획에 사용된다.
- MCTS

### 모델 프리

- 환경의 모델을 사용하지 않고 경험을 통해 직접학습한다.
- DQN

## ON-Policy, off-Policy

### ON-Policy

- 학습 중에 사용하는 정책과 실제로 행동하는 정책이 동일할때
- SARSA

### off-Policy

- 학습 중에 사용하는 정책과 실제로 행동하는 정책이 다를때
- Q-learning