Reinforce Learning

Texonomy & MDP

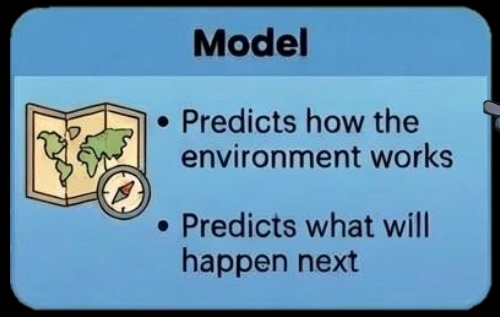
소프트웨어 꼰대 강의

노기섭 교수

(kafa46@hongik.ac.kr)

Types of RL

Texonomy starts with Model



RL Agent

Model-free RL

Agent doesn't know & access environment

Almost of Deep RL cases

Need many samples

Model-based RL

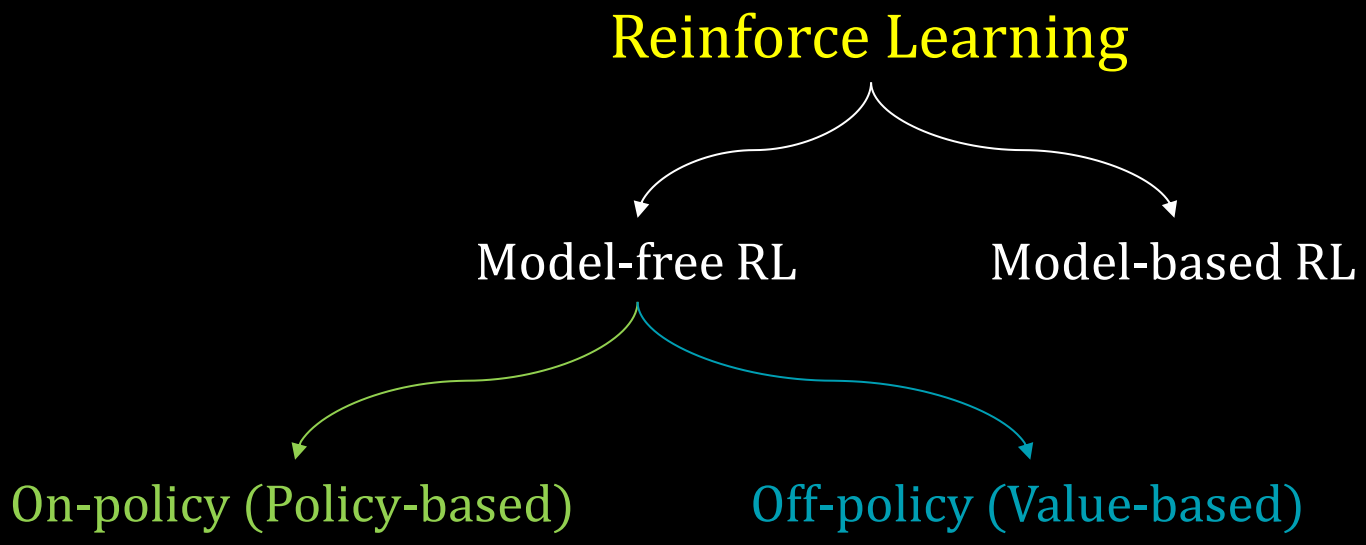
Agent knows environment.

Agent learns the model of environment, then plan using the model

Update & replan the model

Sample efficient (AlphaGo, AlphaZero, ...)

Details in Model-free RL



A Taxonomy of RL Algorithms

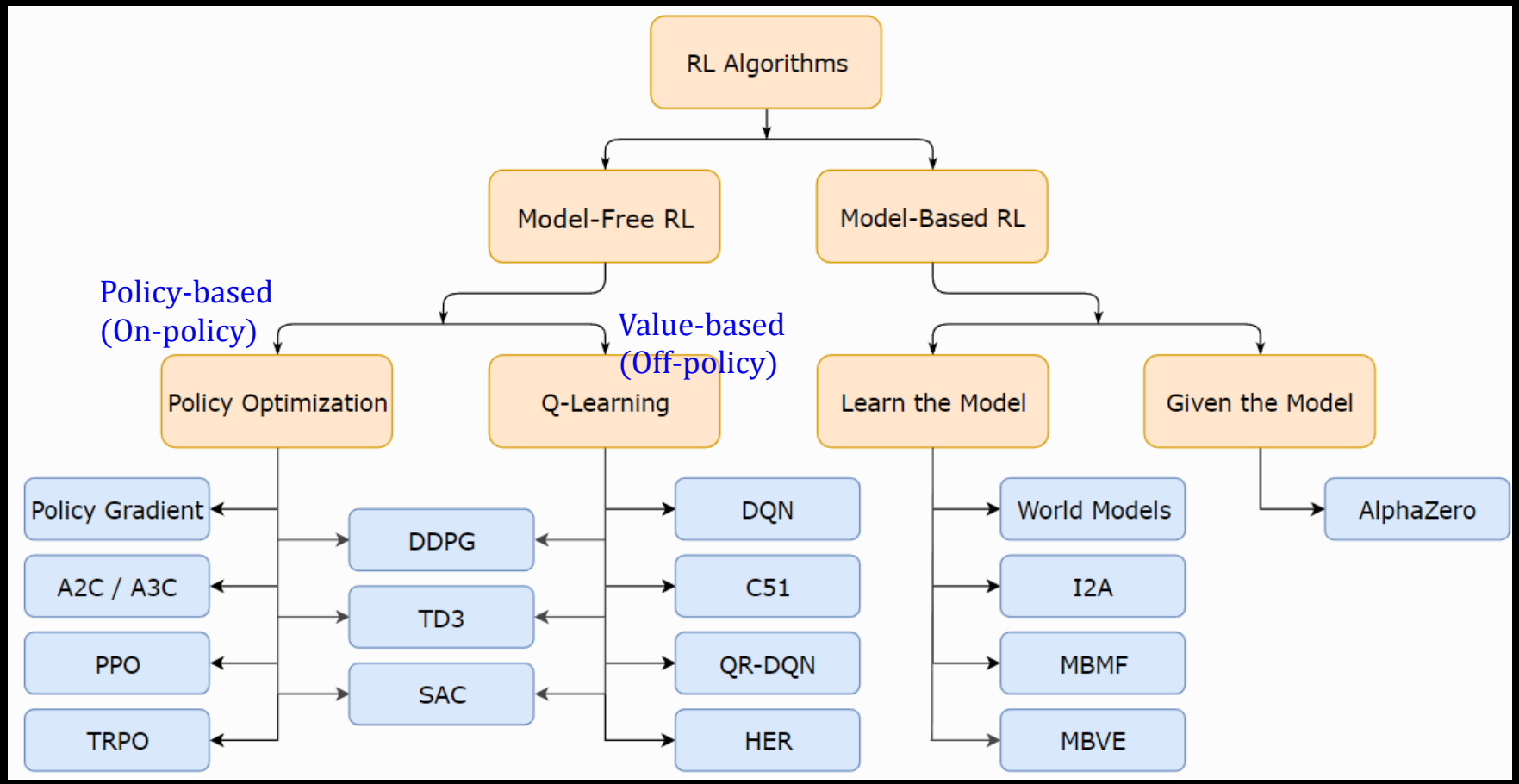


Image source: https://spinningup.openai.com/en/latest/spinningup/rl_intro2.html#citations-below

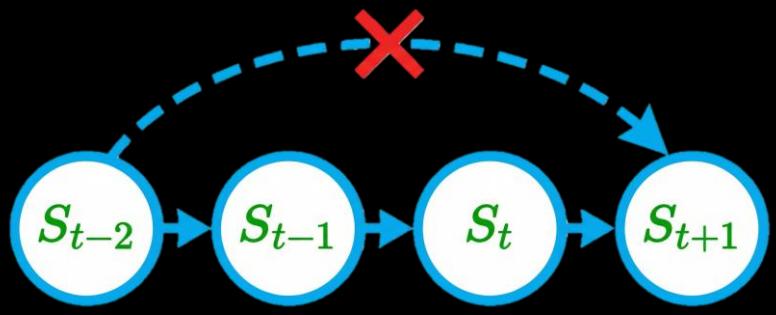
The Markov Property in Reinforcement Learning

Mathematical Definition:

$$P(s_{t+1} \mid s_0, s_1, \cdots, s_t) = P(s_{t+1} \mid s_t)$$

- The future state depends only on the current state, not on the sequence of past states.
- The current state is a sufficient statistic of the history.

The "Memoryless" Property



Why It Matters for Reinforcement Learning?

- Enables local recursion → tractable planning and learning algorithms
- Makes dynamic programming possible (value iteration, policy iteration)
- Foundation for TD methods and model-free/value-based RL approaches

When the Markov Property Breaks

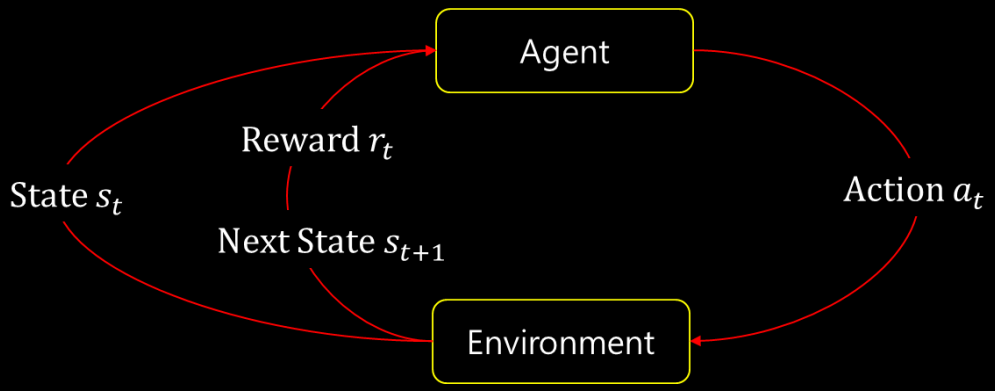
Property is Compromised?

Partial observability:

If need memory of history

- Recurrent Neural Network (RNN)
- Long Short-Term Memory (LSTM)
- Partially Observable Markov Decision Process (POMDP)

Recap: MDP Components

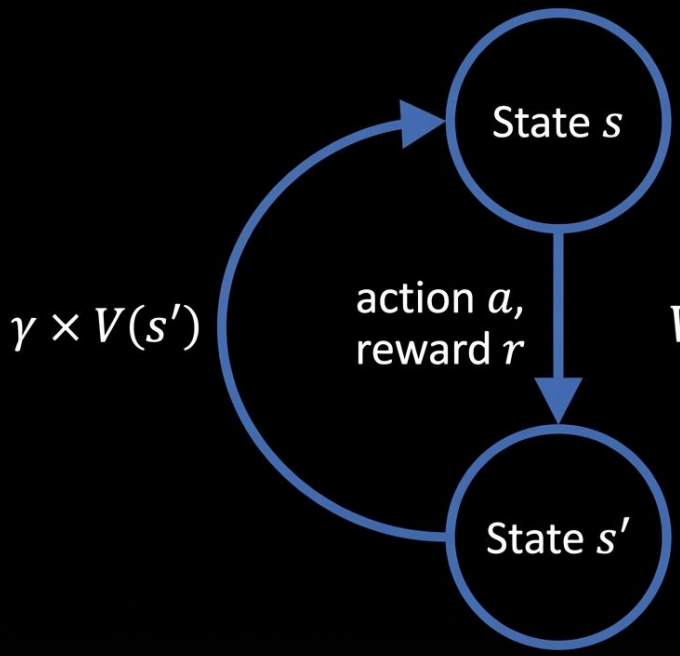


MDP defines the problem!

RL algorithms search for optimal behavior!

마코프 프로세스는 '게임의 규칙',
강화학습은 게임에서 이기기 위한 '전략'

Memoryless Property: Key Foundation in RL



The "Memoryless" Property

$$P(s_{t+1} \mid s_0, s_1, \cdots, s_t) = P(s_{t+1} \mid s_t)$$

$$V(s) = r + \gamma \times V(s')$$

State Transition:
At state s, the agent takes action a.
The environment returns:

- Immediate reward r
- Next state s'

This transition is governed by the Markov process dynamics.

수고하셨습니다 ..^^..