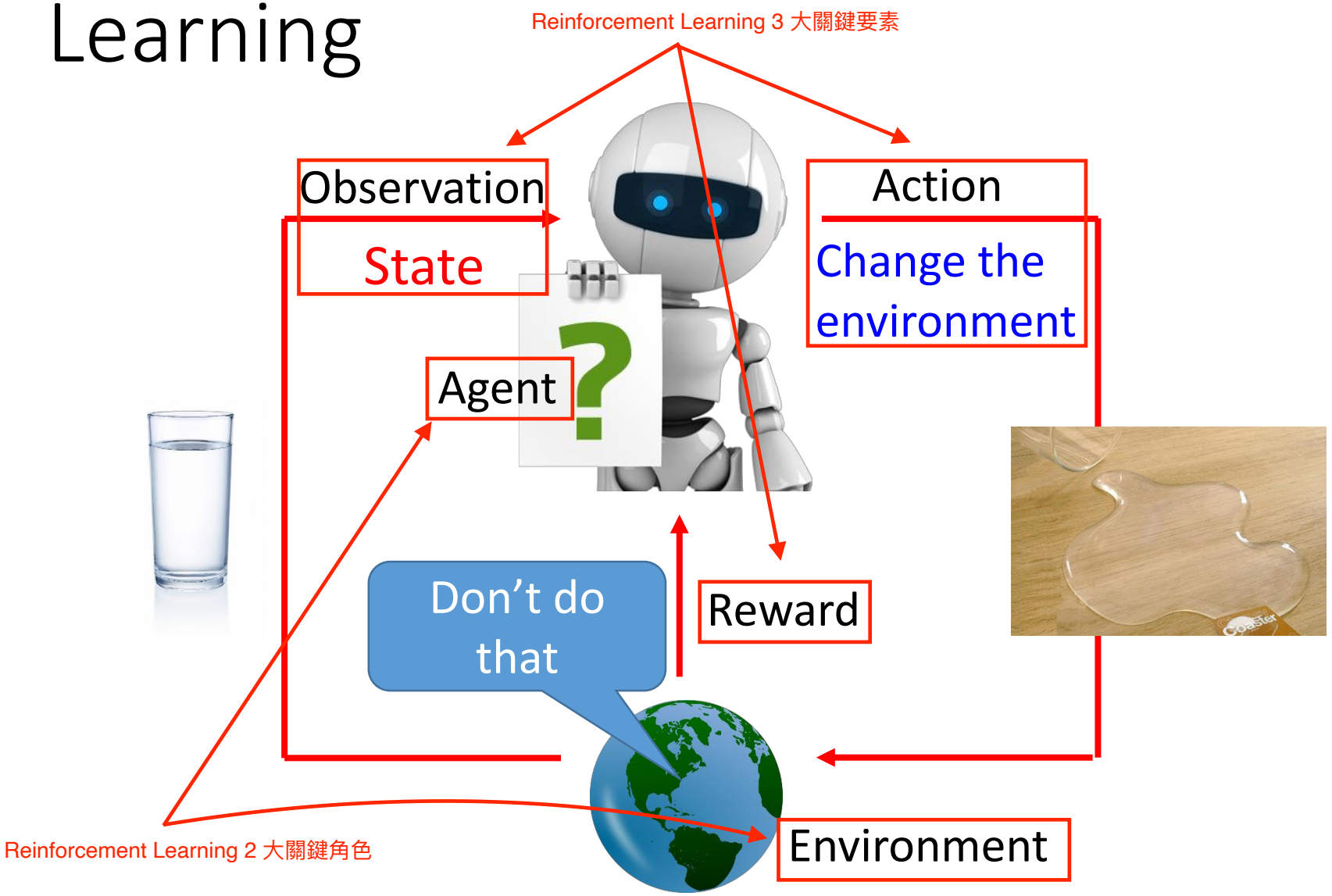
Deep Reinforcement Learning

Scratching the surface

Deep Reinforcement Learning

Deep Reinforcement Learning: AI = RL + DL

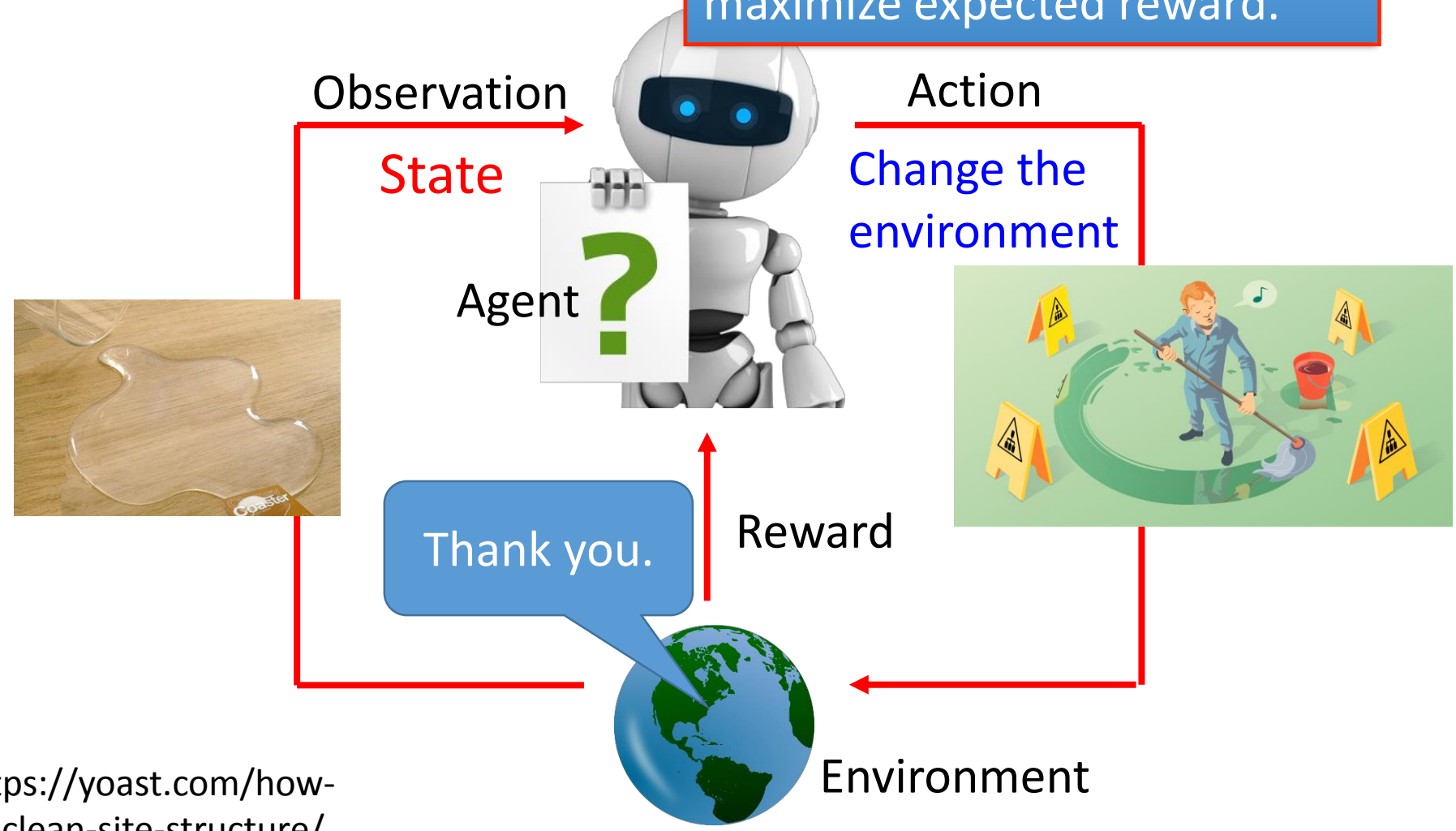
Scenario of Reinforcement Learning



Scenario of Reinforcement Learning

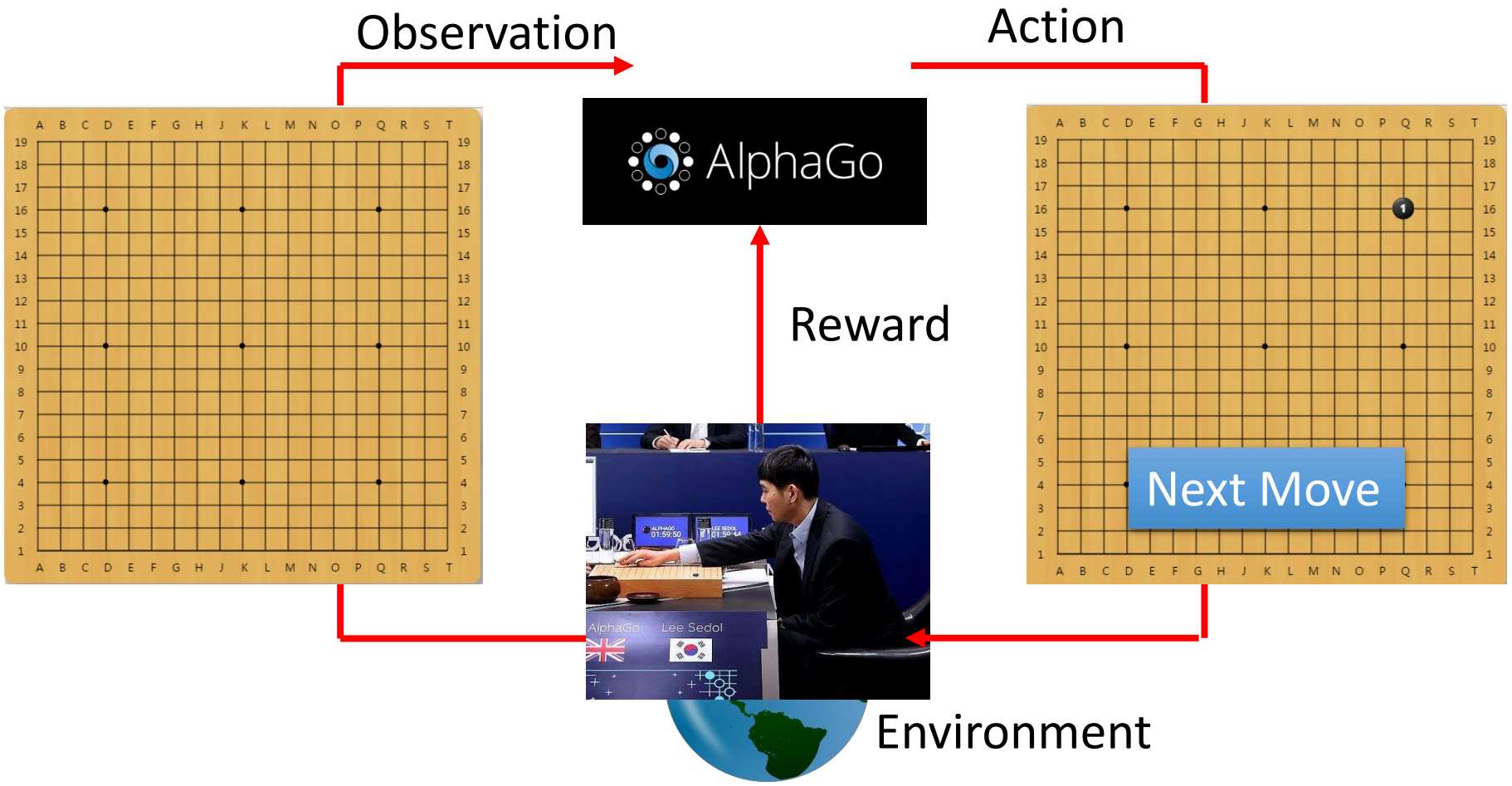
RL 中的 Agent 的唯一目標

Agent learns to take actions to maximize expected reward.



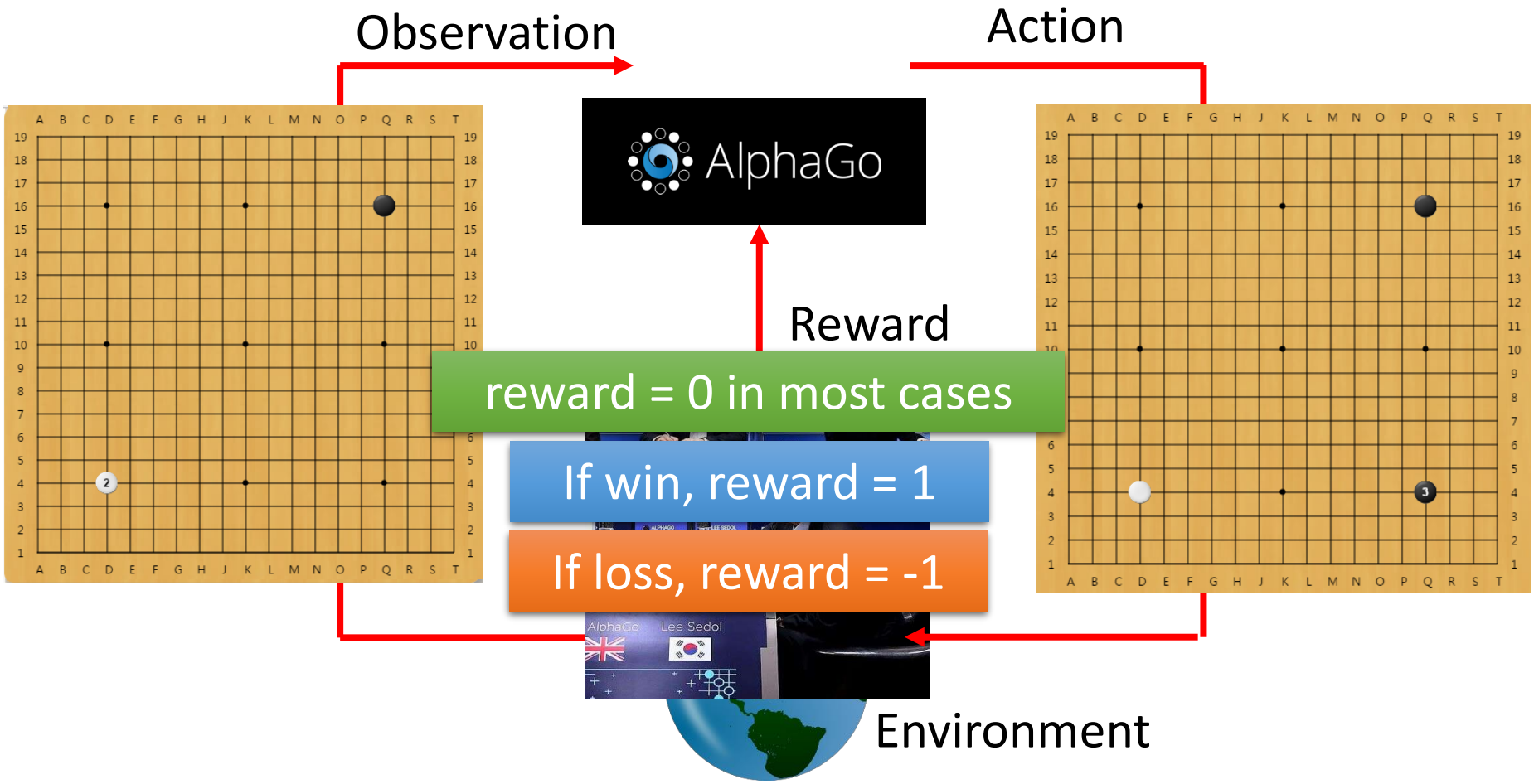
https://yoast.com/how-to-clean-site-structure/

Learning to paly Go



Learning to paly Go

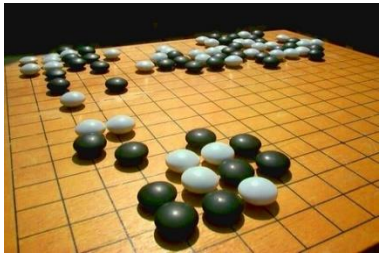
Agent learns to take actions to maximize expected reward.



Learning to paly Go
- Supervised v.s. Reinforcement

Supervised: Learning from teacher

Next move:
"5-5"



Next move:
"3-3"

Reinforcement Learning Learning from experience

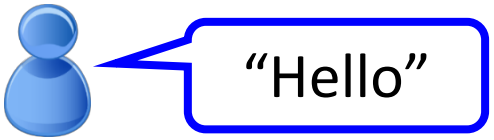
First move → …… many moves …… → Win!

(Two agents play with each other.)

Alpha Go is supervised learning + reinforcement learning.

Learning a chat-bot - Supervised v.s. Reinforcement

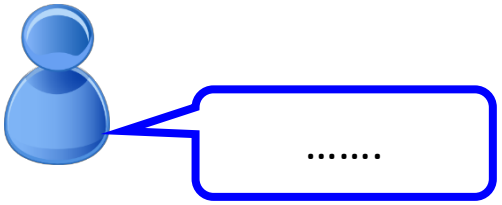
- Supervised

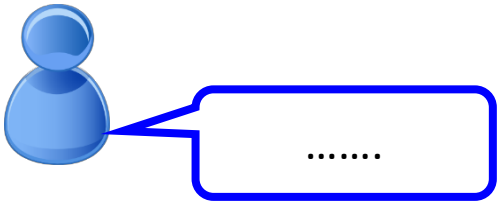


Say “Hi”

Say “Good bye”

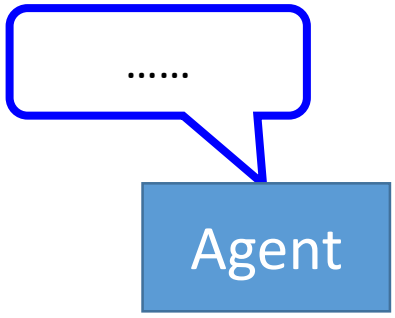
- Reinforcement





.....

Bad



Learning a chat-bot - Reinforcement Learning

- Let two agents talk to each other (sometimes generate good dialogue, sometimes bad)

How old are you?

See you.

See you.

See you.

How old are you?

I am 16.

I though you were 12.

What make you think so?

Learning a chat-bot - Reinforcement Learning

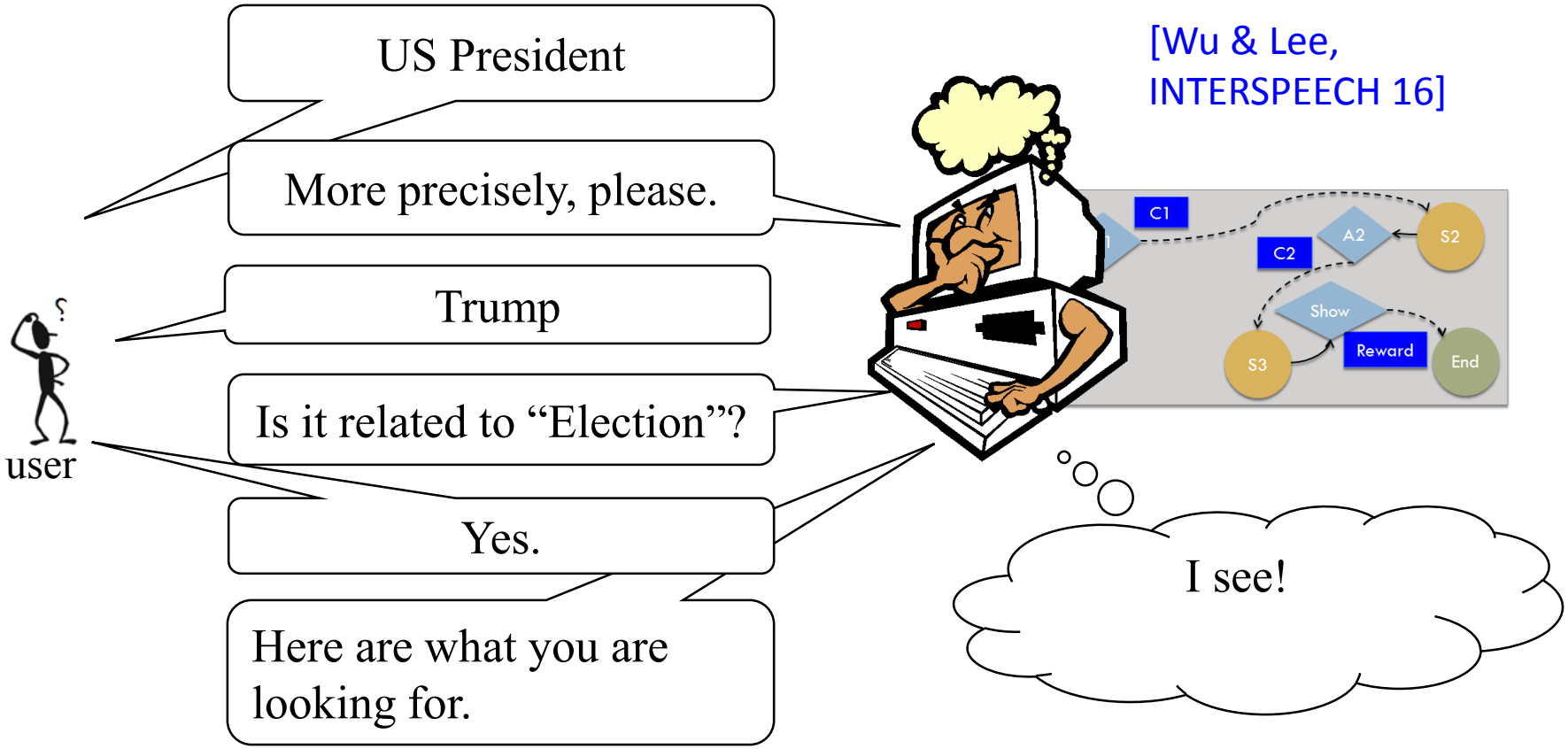
- By this approach, we can generate a lot of dialogues.
- Use some pre-defined rules to evaluate the goodness of a dialogue

Machine learns from the evaluation

Deep Reinforcement Learning for Dialogue Generation
https://arxiv.org/pdf/1606.01541v3.pdf

More applications

- Interactive retrieval



More applications

- Flying Helicopter
  - https://www.youtube.com/watch?v=0JL04JJjocc
- Driving
  - https://www.youtube.com/watch?v=0xo1Ldx3L5Q
- Google Cuts Its Giant Electricity Bill With DeepMind-Powered AI
  - http://www.bloomberg.com/news/articles/2016-07-19/google-cuts-its-giant-electricity-bill-with-deepmind-powered-ai
- Text generation
  - Hongyu Guo, “Generating Text with Deep Reinforcement Learning”, NIPS, 2015
  - Marc'Aurelio Ranzato, Sumit Chopra, Michael Auli, Wojciech Zaremba, “Sequence Level Training with Recurrent Neural Networks”, ICLR, 2016

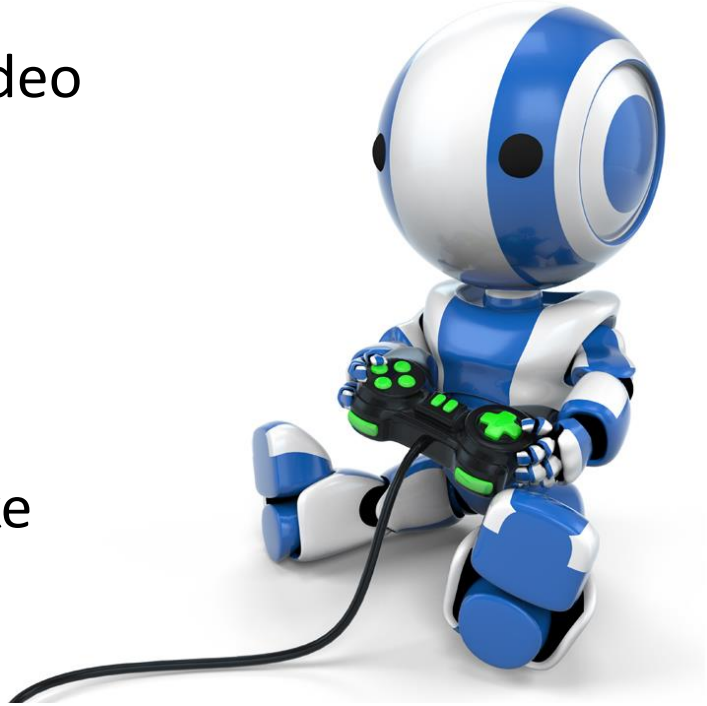
Example: Playing Video Game

現成 Environment

- Widely studies:
  - Gym: https://gym.openai.com/
  - Universe: https://openai.com/blog/universe/

Machine learns to play video games as human players

- What machine observes is pixels
- Machine learns to take proper action itself

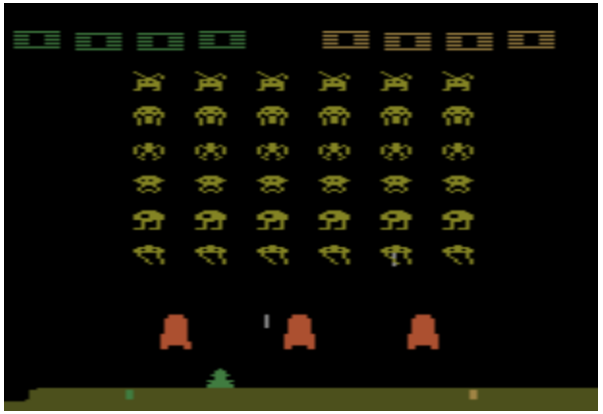


Example: Playing Video Game

Start with observation s_1

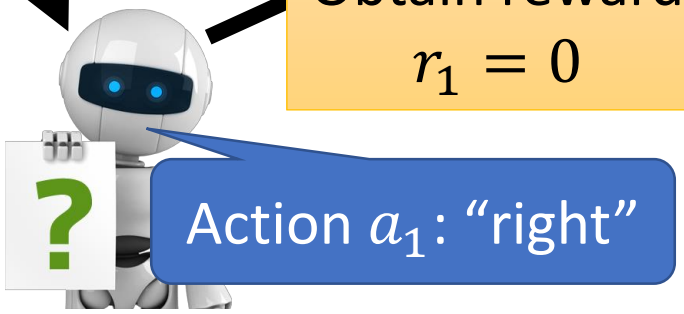
Observation s_2

Observation s_3

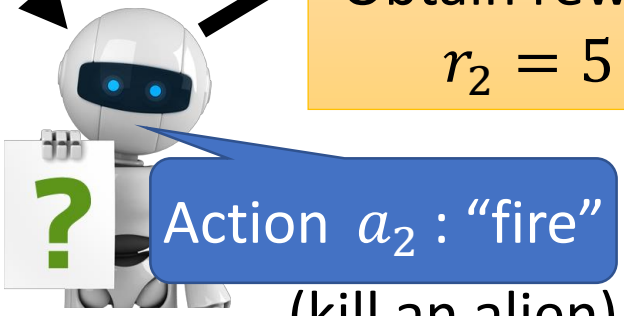


Obtain reward $r_1 = 0$

Obtain reward $r_2 = 5$



Action a_1: "right"



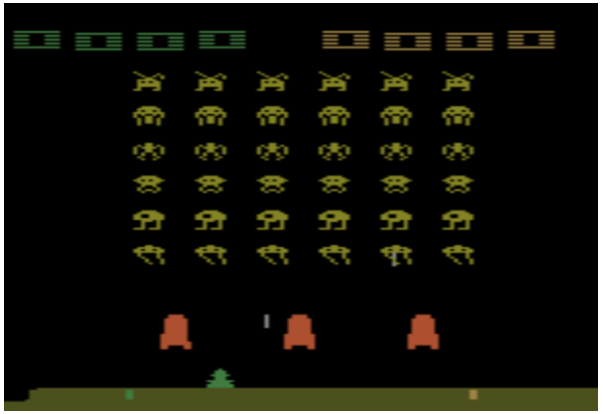
Action a_2 : "fire" (kill an alien)

Usually there is some randomness in the environment

Example: Playing Video Game

Start with observation s_1

Observation s_2



Observation s_3

After many turns

Game Over (spaceship destroyed)

Obtain reward r_T

Action a_T

This is an ***episode***.

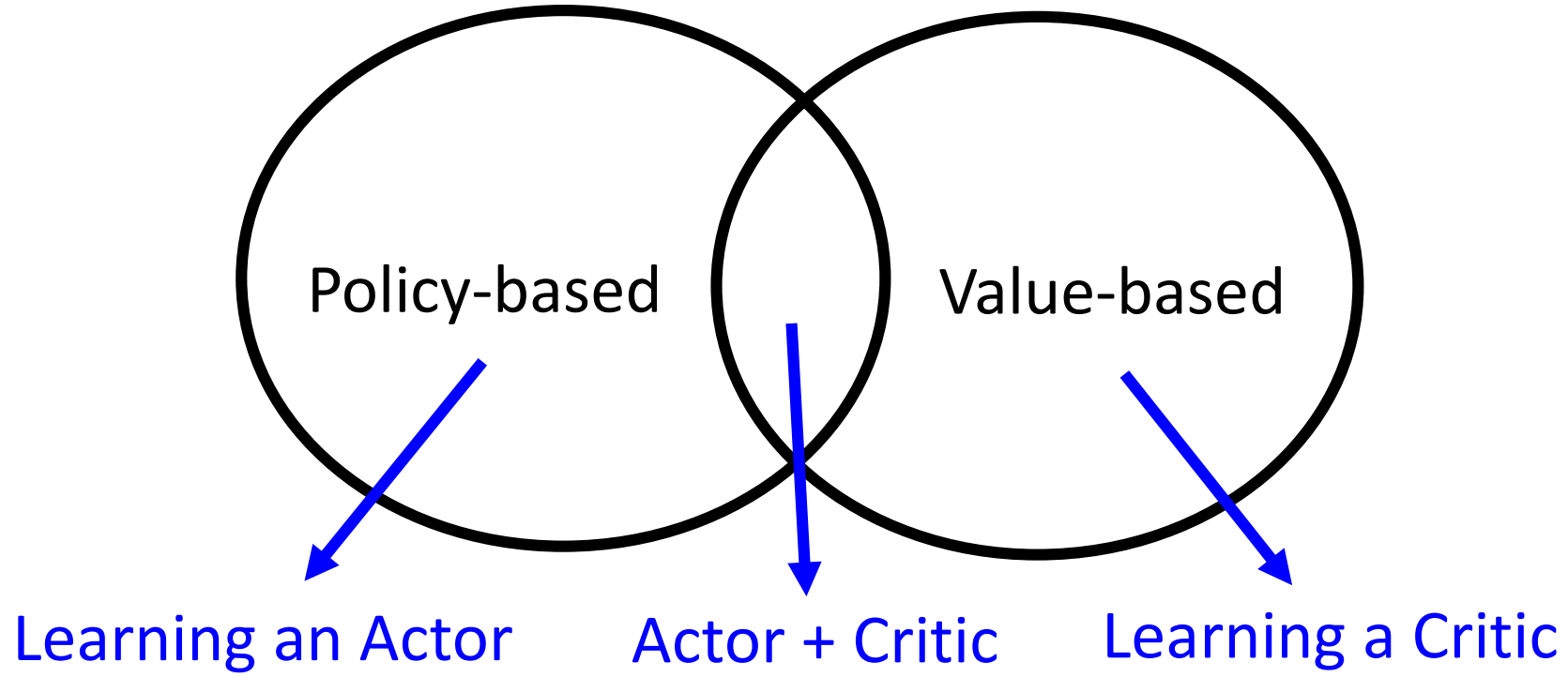
Learn to maximize the expected cumulative reward per episode

Difficulties of Reinforcement Learning

- Reward delay
  - In space invader, only “fire” obtains reward
    - Although the moving before “fire” is important
  - In Go playing, it may be better to sacrifice immediate reward to gain more long-term reward
- Agent’s actions affect the subsequent data it receives
  - E.g. Exploration

Outline

Alpha Go: policy-based + value-based + model-based



Asynchronous Advantage Actor-Critic (A3C)

Volodymyr Mnih, Adrià Puigdomènech Badia, Mehdi Mirza, Alex Graves, Timothy P. Lillicrap, Tim Harley, David Silver, Koray Kavukcuoglu, "Asynchronous Methods for Deep Reinforcement Learning", ICML, 2016