Deep Reinforcement Learning for deep learning experts

Vikas Dhiman

October 3, 2022

Prerequisites for knowing Reinforcement Learning

1. Linear algebra
2. Probability
3. Python

Prerequisites for knowing Deep Reinforcement Learning

1. Deep Learning

BF Skinner's Reinforcement Learning for Pigeons

Video

[1]

[1] Image source:bfskinner.org

DRL for DL experts

BF Skinner's Reinforcement Learning for Pigeons

BF Skinner's Reinforcement Learning for Pigeons

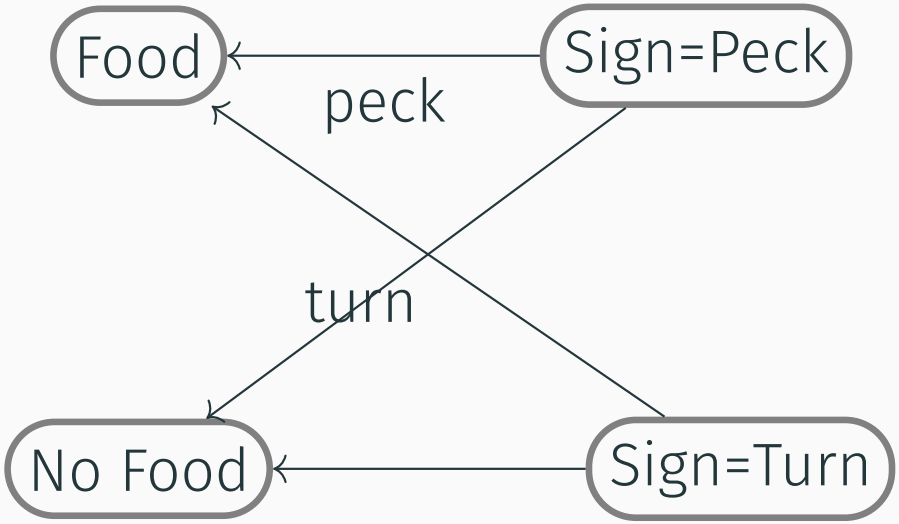
Video

[1]

[1] Image source:bfskinner.org

1. BF Skinner demonstrated that pigeons could learn to repeat an action that lead them to a particular reward. [1, p15]

RL terminology



State ($\mathbf{s}_t \in \mathcal{S}$) Example: Sign is peck or turn. Food is dispensed or not.

Reward function ($r_t(\mathbf{s}_t) \to \mathbb{R}$) Example: Food is high reward ($r_t = 100$). food is zero-reward ($r_t = 0$).

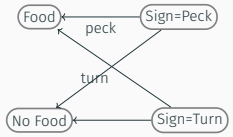
Actions ($\mathbf{a}_t \in \mathcal{A}$) Example: To peck or to turn or no action.

Transition probabilities ($T(\mathbf{s}_{t+1}|\mathbf{s}_t, \mathbf{a}_t) \to [0, 1]$) Example: Probability of food dispensing if you peck when Sign-peck is shown.

DRL for DL experts

RL terminology

RL terminology



State ($\mathbf{s}_t \in \mathcal{S}$) Example: Sign is peck or turn. Food is dispensed or not.

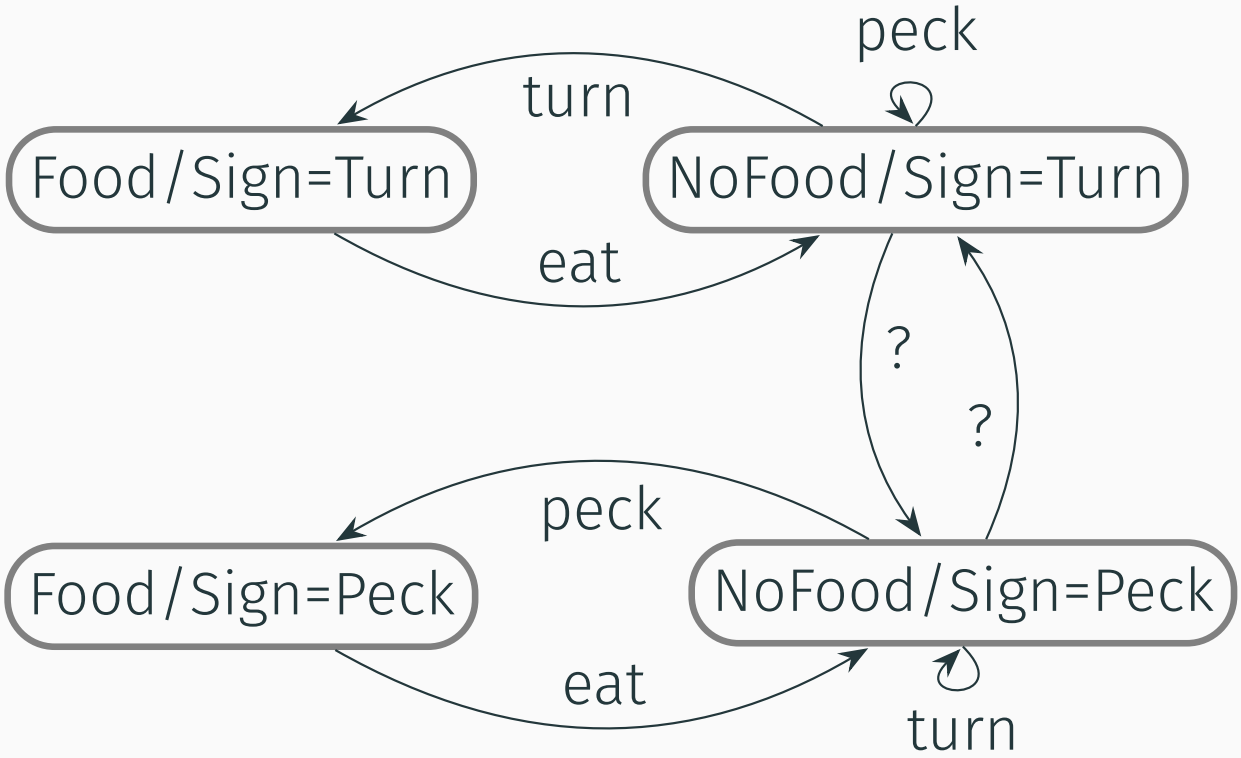
Reward function ($r_t(\mathbf{s}_t) \to \mathbb{R}$) Example: Food is high reward ($r_t = 100$). food is zero-reward ($r_t = 0$).

Actions ($\mathbf{a}_t \in \mathcal{A}$) Example: To peck or to turn or no action.

Transition probabilities ($T(\mathbf{s}_{t+1}|\mathbf{s}_t, \mathbf{a}_t) \to [0, 1]$) Example: Probability of food dispensing if you peck when Sign-peck is shown.

State is the full description of the world at time t that captures the entire history. Example: in this example the state can be captured with two bits $\mathbf{s}_t = [f_t; p_t]$, where $f_t \in \{0, 1\}$ describes a food or no food state and $p_t \in \{0, 1\}$ describes the sign showing peck or turn.

Better state diagram



RL problem

Policy function $\pi(\mathbf{s}_t) \rightarrow \mathbf{a}_t$

Discount factor $\gamma \in (0, 1)$.

$$\pi^*(.) = \arg \max_{\pi} \mathbb{E}_T \left[\sum_{t=0}^{\infty} \gamma^t r(\mathbf{s}_t) \right]$$

$$\text{such that } \mathbf{s}_{t+1} \sim T(.|\mathbf{s}_t, \pi(\mathbf{s}_t)) \forall t \in [k, \infty)$$

$$\text{and } \mathbf{s}_0 \sim p_0(.)$$

Value Function

$$\pi^*(.) = \arg \max_{\pi} \mathbb{E}_T \left[\sum_{t=0}^{\infty} \gamma^t r(\mathbf{s}_t) \right]$$

$$\text{such that } \mathbf{s}_{t+1} \sim T(.|\mathbf{s}_t, \pi(\mathbf{s}_t)) \forall t \in [k, \infty)$$

$$\text{and } \mathbf{s}_0 \sim p_0(.)$$

$$V_\pi(\mathbf{s}_k) = \mathbb{E}_T \left[\sum_{t=k}^{\infty} \gamma^t r(\mathbf{s}_t) \right]$$

$$\text{such that } \mathbf{s}_{t+1} \sim T(.|\mathbf{s}_t, \pi(\mathbf{s}_t)) \forall t \in [k, \infty)$$

Action Value Function

$$Q_\pi(\mathbf{s}_k, \mathbf{a}_k) = \mathbb{E}_T \left[\sum_{t=k+1}^{\infty} \gamma^t r(\mathbf{s}_t) \right]$$

such that $\mathbf{s}_{t+1} \sim T(.|\mathbf{s}_t, \pi(\mathbf{s}_t)) \forall t \in [k, \infty)$

References


Richard S Sutton and Andrew G Barto.
Reinforcement learning: An introduction.
MIT press, 2020.