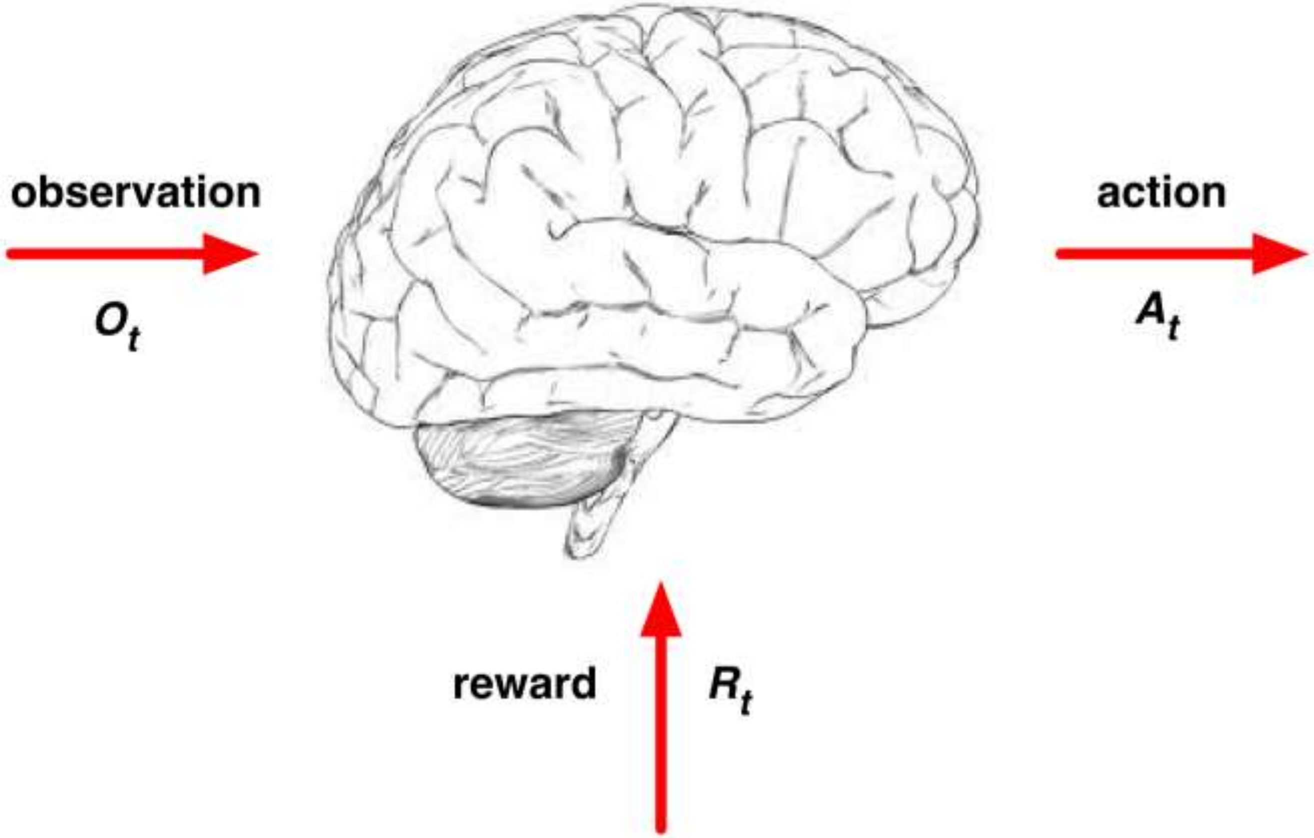
Reinforcement Learning

主讲人：沈楚云

Agent and Environment



History and State

- The history is the sequence of observations, actions, rewards
- Ht = O1, R1, A1, ..., At−1, Ot, Rt
- State is the information used to determine what happens next
- state is a function of the history: St = f(Ht)

Markov State

- A state St is Markov if and only if

  $P[St+1 \mid St] = P[St+1 \mid S1, ..., St]$

- Once the state is known, the history may be thrown away

Differences between RL and other ML method

- **supervised learning**
    - There is no supervisor, only a reward signal.
    - Feedback is delayed, not instantaneous.
    - Agent’s actions affect the subsequent data it receives.
    - Time really matters (sequential, non i.i.d data).

- **unsupervised learning**
    - Reinforcement learning is also different from what machine learning researchers call unsupervised learning, which is typically about finding structure hidden in collections of unlabeled data.

Importance components of an RL agent

- An RL agent may include one or more of these components:
  - Policy: agent’s behavior
  - function Value function: how good is each state and/or action
  - Model: agent’s representation of the environment

Policy

- A policy is the agent’s behavior
- Deterministic policy: $a = \pi(s)$
- Stochastic policy: $\pi(a|s) = P[At = a|St = s]$

Value Function

- Value function is a prediction of future reward Used to evaluate the goodness/badness of states And therefore to select between actions, e.g.
- $V\pi(s) = E\pi[Rt+1 + \gamma Rt+2 + \gamma^2Rt+3 + \ldots \mid St = s]$

Model

- A model predicts what the environment will do next
- P predicts the next state
- R predicts the next (immediate) reward, e.g.

$$\mathcal{P}_{ss'}^{a} = \mathbb{P}[S_{t+1} = s' \mid S_t = s, A_t = a]$$

$$\mathcal{R}_{s}^{a} = \mathbb{E}\left[R_{t+1} \mid S_t = s, A_t = a\right]$$

Categorizing RL agents

- Value Based
  - No Policy (Implicit)
  - Value Function
- Policy Based
  - Policy
  - No Value Function
- Actor Critic
  - Policy
  - Value Function