SCIENCE HUB FOR HUMANITY
AND ARTIFICIAL INTELLIGENCE
UCLA amazon

UCLA Samueli
Electrical & Computer Engineering

Learning in safety-critical, multi-agent, and lifelong systems: Bandits and RL approaches

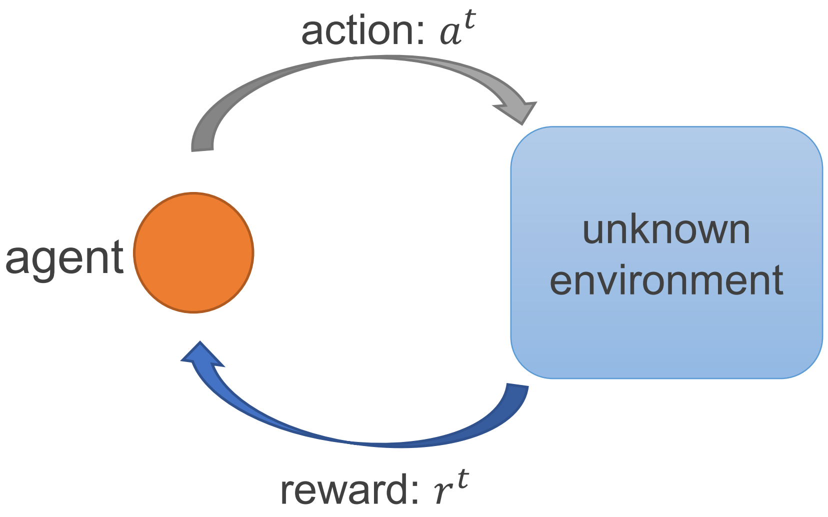
Sanae Amani
Advisor: Prof. Lin Yang

02/23/2023

Background

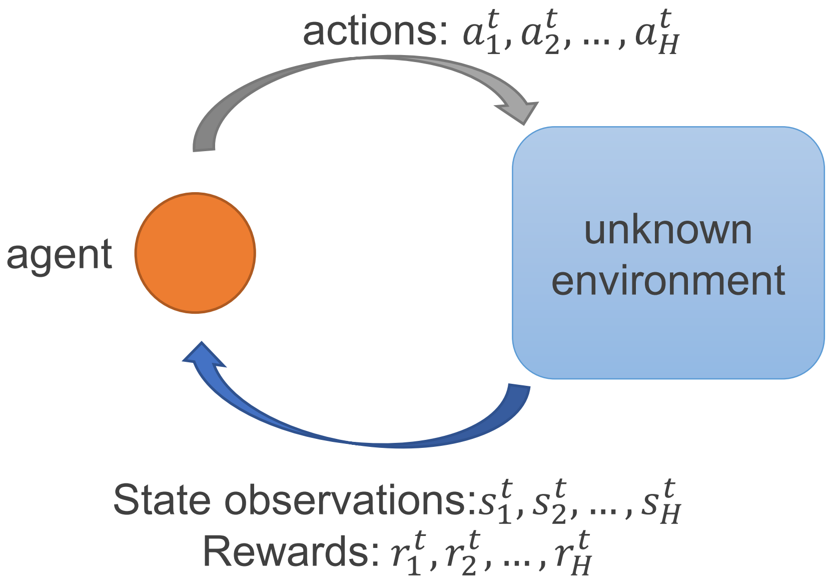
Multi-armed Bandit (MAB):

a single-stage interactive learning framework



Reinforcement Learning (RL):

a multi-stage interactive learning framework



Multi-armed Bandit

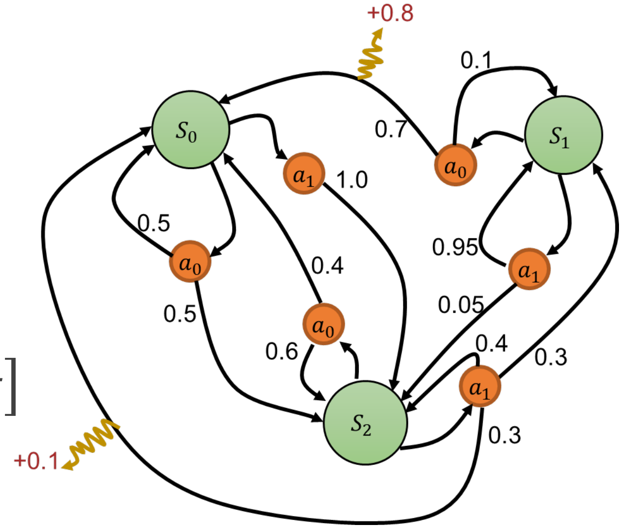
Multi-armed Bandit (MAB): a single-stage interactive learning framework

- Action set A
- Reward $r(a)$.
- Goal (without knowledge of r): $\underset{a \in A}{\text{maximize}}\ r(a)$

Reinforcement Learning

Reinforcement Learning (RL): a multi-stage interactive learning framework

- MDP: $M \coloneqq (S, A, H, P, r)$
- Transition kernel $P(s'|s, a)$, reward $r(s, a)$.
- Horizon H.
- A policy $\pi: S \to \Delta(A)$
- Goal (without knowledge of P and r): $\underset{\pi}{\text{maximize}}\ V^{\pi} \coloneqq E\left[\sum_{h=1}^{H} r_h \middle|\, \pi\right]$



When the model is known, solve by dynamic programming.

Motivation 1

❖ **Safety-critical systems:**

Self-driving cars

Medical trials

❖ **Challenges:**

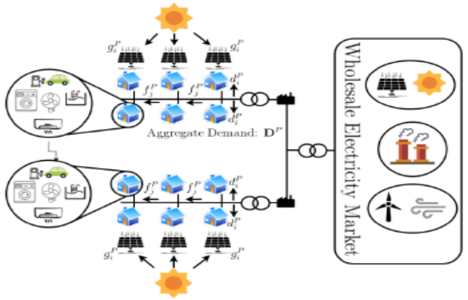
- Playing unsafe actions/policies may result in catastrophic results
- Safety requirements are typically unknown and must be learned.

Our research goal: being safe while achieving good performance comparable to unsafe approaches

Motivation 2

- **Multi-Agent systems:**

Robotics



Smart energy grid

- **Challenges:**
  - Certain systems are distributed inherently.
  - Distributed solutions speed up the process.

Our research goal: improve communication and performance efficiency over prior work in muti-agent systems

Motivation 3

❖ **Lifelong learning systems:**

Robotics

Self-driving cars

❖ **Challenges:**

- ➢ Learning a multi-task policy while solving a streaming sequence of arbitrary tasks.
- ➢ Computationally efficient solutions

Our research goal: solutions that are provably computationally efficient while achieving good performance comparable to direct extensions of single-task approaches.

Thank you very much!