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

State observations: $s_1^t, s_2^t, ..., s_H^t$
Rewards: $r_1^t, r_2^t, ..., r_H^t$
Multi-armed Bandit

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

• Action set $A$
• Reward $r(a)$.
• Goal (without knowledge of $r$): maximize $\underset{a \in A}{\max} r(a)$
Reinforcement Learning

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

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

When the model is known, solve by dynamic programming.
Motivation 1

- Safety-critical systems:

  - 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:**

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

Our research goal: improve communication and performance efficiency over prior work in multi-agent systems.
Motivation 3

- Lifelong learning systems:

- 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!