Whittle Index & Restless Bandits

Henry Vu

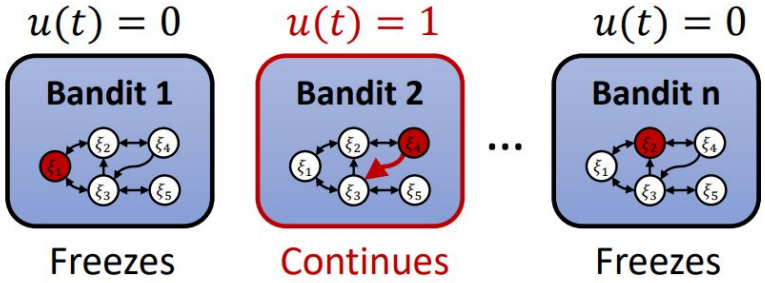
March 29, 2024

Outline

- Restless Bandit Problem
  - Optimization Problem
- Indexability, Whittle Index
- Gittins Index Theorem

Simple Family of Alternative Bandit Processes

- **n Markov Bandit Processes** with state space $S = S_1 \times S_2 \times \ldots \times S_n$.
- Control u(t) = 1 (*active*) is applied to a single bandit i_t at each decision time t. Transition probability $P_{it}(s'|s_{it}(t))$
- Control u(t) = 0 (*passive*) is applied to all other bandits. These bandits remain in the **same state**.



Restless Bandits

- n Markov Bandits. At time t, apply u(t) = 1 to exactly **m bandits** and u(t) = 0 to all other bandits.

- Action u(t) = 0 **no longer freezes** the bandit.
  - They **evolve** (possibly) and **accrue reward** in a *different* way (different Markov chain) than when u = 1.

- Example: work ⟶ more fatigue, rest ⟶ recovery.

- Finding optimal control policy π^* for restless bandits is intractable

Problem Formulation

- Optimization Problem:

$$\begin{aligned} \text{maximize} \quad & \lim_{T\to\infty} \mathbb{E}\left[\sum_{t=0}^{T-1} \gamma^t \sum_{i=1}^{n} r_i(s_i, u_i)\right] \\ \text{subject to} \quad & \sum_{i=1}^{n} u_i(t) = m, \quad \forall t, \\ & u_i(t) \in \{0, 1\}, \quad \forall i. \end{aligned}$$

- Restless MAB is provably hard, PSPACE-complete (Papadimitriou & Tsitsiklis, 1999)

Problem Formulation

- Relaxed activation constraints:
  - Replacing the hard constraint $\sum_{i=1}^{n} u_i(t) = m$ by the “time-averaged constraint”.
  - At each time t, the number of active arms is m only in **expectation**. The hard constraint is $\Sigma_i u_i(t) = m \ \forall t$, and the relaxed constraint is $E[\Sigma_i u_i(t)] = m$.
  - Incorporating discount factor γ, we have the following relaxed constraint

$$\mathbb{E}\sum_{t=0}^{\infty}\gamma^t\sum_{i=1}^{n}u_i(t) = \sum_{t=0}^{\infty}\gamma^t m = m/(1-\gamma)$$

Relaxed Formulation

- Relaxed:

$$\begin{aligned} \text{maximize} \quad & \lim_{T\to\infty} \mathbb{E}\left[\sum_{t=0}^{T-1}\gamma^t \sum_{i=1}^{n} r_i(s_i,u_i)\right] \\ \text{subject to} \quad & \mathbb{E}\sum_{t=0}^{T-1}\gamma^t \sum_{i=1}^{n} u_i(t) = m/(1-\gamma), \\ & u_i(t)\in\{0,1\}, \quad \forall i. \end{aligned}$$

- **Lagrange function** is given by:

$$\text{maximize} \lim_{T\to\infty} \mathbb{E}\left[\sum_{t=0}^{T-1}\gamma^t \sum_{i=1}^{n} \left(r_i(s_i,u_i) - \lambda u_i(t)\right)\right] + \lambda\left(m/(1-\gamma)\right)$$

Decoupled Problem

- Notice that we can neglect the last term (constant) and **decouple** this problem. Then, for a fixed $\lambda \geq 0$ and for each bandit, we have:

$$\text{maximize} \lim_{T \to \infty} \mathbb{E} \left[\sum_{t=0}^{T-1} \gamma^t \left(r_i(s_i, u_i) - \lambda u_i(t) \right) \right]$$

- Similar to Gittins Index!

Solution to the Decoupled Problem

- Main difference when compared to Gittins Index is that passive bandits may **change state** and **accrue reward**. Thus, the optimal policy for the Decoupled Problem may NOT be a **stopping rule**.

- In general, the optimal policy divides the state space into two subsets:
  - Let $P(\lambda)$ be the set of states for which it is **optimal to freeze** when the playing charge is λ (similar to Gittins).
  - $u(t) = 1$ if $s(t) \in P^C(\lambda)$; stop otherwise.

Indexability

- Definition of **Indexability**: The Decoupled Problem associated with bandit i is indexable if P(λ) increases monotonically from ∅ to the entire state space as λ increases from -∞ to +∞. The restless MAB problem is **indexable** if the Decoupled Problem is **indexable for all bandits**.

Thank you!