Ming Hsieh Department of Electrical Engineering

USC Viterbi School of Engineering

Tracking a Stochastic Sequence

Parisa Mansourifard, Bhaskar Krishnamachari, in Collaboration with Tara Javidi

Introduction

- We want to track a stochastic sequence, $B(t)$
- The states $i = 1,...,M$
- Transition matrix, P
- Partially Observable Markov Decision Process
- Finite horizon T, time steps $t = 1,...,T$
- Our belief vector at time t,

$$b_t = [b_t(1),...,b_t(M)]$$

$$b_t(i) = \Pr(B(t) = i), i = 1,...,M$$

- Choose an action, a, if higher than B(t), we will find out the exact amount of B(t). Otherwise, we can make our belief narrower.
- Evolution of belief vector :

$$b_{t+1} = \begin{cases} T_a b_t P, & if\ a \le B(t) \\ I_{B(t)} P, & if\ a > B(t) \end{cases}$$

$$T_a b(i) = \begin{cases} 0, & if\ i < a \\ \dfrac{b(i)}{\sum_{j=r}^{M} b(j)}, & if\ i \ge a \end{cases}$$

- Current reward :

$$R(B(t);a) = \min(a, B(t)) - C(a - B(t))^+$$

Dynamic Programming

$$V_t(b_t) = \max_{a=1,...,M} V_t(b_t;a), \quad \forall t = 1,...,T$$

$$V_T(b_T;a) = \overline{R}(b_T;a),$$

$$V_t(b_t;a) = \overline{R}(b_t;a) + \beta V_t^{\,f}(b_t;a)$$

$$V_t^{\,f}(b_t;a) = E\{V_{t+1}(b_{t+1}) \mid a\}$$

$$= \sum_{i=a}^{M} b_t(i) V_{t+1}(T_a b_t P) + \sum_{i=1}^{a-1} b_t(i) V_{t+1}(I_i P)$$

Properties of Value Function

- Convexity with respect to belief vector

$$V_t(\lambda b_1 + (1-\lambda) b_2) \le \lambda V_t(b_1) + (1-\lambda) V_t(b_2)$$

$$0 \le \lambda \le 1$$

- Monotonically increase of future expected reward

$$V_t^{\,f}(b_t;a_1) \ge V_t^{\,f}(b_t;a_2),$$

$$a_1 \ge a_2$$

Myopic Policy

ignoring the impact of the current action on the future reward, myopic policy is given by

$$a_t^{\,Myopic}(b) = \arg\max_{a=1,...,M} \overline{R}(b_t;a)$$

$$= \min\{a = 1,...M \mid \sum_{i=1}^{a} b_t(i) \ge \frac{1}{1+C}\}$$

Problem Formulation

- Policy vector: $\pi = [\pi(1),...,\pi(T)]$
- Selecting an action $\pi(t) = a_t \in \{1,...,M\}$
- Maximizing total discounted expected reward:

$$\max_{\pi} E^{\pi}[\sum_{t=1}^{T} \beta^{t-1} R(b_t;a_t) \mid b_1]$$

- defining value function, i.e. maximum expected remaining reward starting from time t: $V_t(b)$

Optimal Policy

Our goal is to find the optimal policy or prove that it has a threshold structure.

$$a_t^{\,Optimal} = \arg\max_{a=1,...,M} V_t(b_t;a)$$

Bounds on Optimal Action:

$$a_t^{\,Myopic} \le a_t^{\,Optimal} \le \max\{i = 1,...,M \mid b_t(i) \ne 0\}$$

Future work: Find a tighter upper bound

parisama@usc.edu