STATS 710 – Seq. Dec. Making with mHealth Applications Dec 06, 2016

Lecture 23: Contextual Decision Processes with Low Bellman Rank are PAC-Learnable

Guest Lecturer: Nan Jiang *Scribe: Hyesun Yoo*

1 Contexual Decision Processes

This lecture was given by Nan Jiang and he discussed about his recent paper [JKA+16].
Let context space $\mathcal{X}$, action space $\mathcal{A}$, horizon H
In each episode,

$$\begin{cases} x_1 \in \mathcal{X} \text{ is drawn, play } a_1 \in \mathcal{A} \\ r_1, x_2 \text{ are drawn, play } a_2 \\ \vdots \\ r_{H-1}, x_H \text{ are drawn, play } a_H \\ r_H \text{ is drawn. End} \end{cases}$$

Policy $\pi : \mathcal{X} \to \mathcal{A}$.
Value $V^\pi = \mathbb{E}\left[\sum_{h=1}^H r_h | a_{1:H} \sim \pi\right]$

2 Value-based RL with function approximation

Given $\mathcal{F} \subseteq (\mathcal{X} \times \mathcal{A} \to [0,1])$ assume $|\mathcal{F}|=$ N$< \infty$
We want to identify $f \in \mathcal{F}$ such that

- V^{π_f} is high, where $\pi_f : x \mapsto \text{argmax}_{a \in \mathcal{A}} f(x,a)$
- f obeys a set of Bellman equation (or "f is valid" for short): $\mathcal{E}(f, \pi_{f'}, h)$=0, $\forall f' \in \mathcal{F}$, $h \in [H]$

where Bellman Error is

$$\mathcal{E}(f, \pi, h) := \mathbb{E}\left[f(x_h, a_h) - r_h - f(x_{h+1}, a_{h+1}) | a_{1:h-1} \sim \pi, a_{h:h+1} \sim \pi_f\right]$$

Formally, we aim to achieve value $V^\star_{\mathcal{F}} = \sup_{f \in \mathcal{F}:\text{ f is valid}} V^{\pi_f}$

3 Bellman rank

Define the Bellman error matrix for each level h as an $N \times N$ matrix, with the (f', f)-th entry being $\mathcal{E}(f, \pi_{f'}, h)$. Informally, Bellman rank M is the maximal rank of such matrices over all h, and is naturally small for a number of RL settings.

PAC-learning goal: identify $f \in \mathcal{F}$ such that $V^{\pi_f} \geq V^\star_{\mathcal{F}} - \epsilon$ with probablity at least $1-\delta$, using only poly$(M, |\mathcal{A}|, H, \log(N/\delta), 1/\epsilon)$ episodes of data.

4 Algorithm: OLIVE

Here is a simplified version of the algorithm OLIVE. This will be discussed in next lecture.

1. Let $V_f = \mathbb{E}\left[\max_{a \in \mathcal{A}} f(x, a)\right]$
2. Let $\mathcal{F}_0 = \mathcal{F}$
3. For every epoch t=1,2,...
4. Pick $f_t = \operatorname{argmax}_{f \in \mathcal{F}_{t-1}} V_f$ and $\pi_t = \pi_{f_t}$.
5. If $\sum_{h=1}^{H} \mathcal{E}(f_t, \pi_t, h) \leq \varepsilon$
   Terminate and return π_t
   End if
6. Pick $h_t \in [1, \cdots, H]$ such that $\mathcal{E}(f_t, \pi_t, h_t) \geq \frac{\varepsilon}{H}$
7. Collect data $\{(x_1^{(i)}, a_1^{(i)}, r_1^{(i)}, \cdots, x_{h_t}^{(i)}, a_{h_t}^{(i)}, r_{h_t}^{(i)}, x_{h_t+1}^{(i)})\}_{i=1}^{n}$
8. $\forall f \in \mathcal{F}_{t-1}$, let $\hat{\mathcal{E}}(f, \pi_t, h_t) = \frac{1}{n} \sum_{i=1}^{n} \frac{I(a_{h_t}^{(i)} = \pi_f(x_{h_t}^{(i)}))}{1/|\mathcal{A}|} \left(f(x_{h_t}^{(i)}, a_{h_t}^{(i)}) - r_{h_t}^{(i)} - f(x_{h_t+1}^{(i)}, \pi_f(x_{h_t+1}^{(i)})) \right)$
9. Let $\mathcal{F}_t := \{f \in \mathcal{F}_{t-1} : |\hat{\mathcal{E}}(f, \pi_t, h_t)| \leq \phi\}$
10. End for

References

[JKA+16] Nan Jiang, Akshay Krishnamurthy, Alekh Agarwal, John Langford, and Robert E Schapire. Contextual decision processes with low bellman rank are PAC-learnable. *arXiv preprint arXiv:1610.09512*, 2016.