STATS 710 – Seq. Dec. Making with mHealth Applications Dec 8, 2016

Lecture 24: The OLIVE Algorithm

Instructors: Susan Murphy and Ambuj Tewari *Scribe: Huajie Qian*

1 Recap: Optimism Led Iterative Value-function Elimination (OLIVE)

1.1 Notation and Assumption

$$K := |\mathcal{A}|, N := |\mathcal{F}|, M := \text{Bellman rank}$$
$$V_f := \mathbb{E}_{x_1}[\max_{a\in\mathcal{A}} f(x_1, a)]$$
$$[H] := \{1, 2, \ldots, H\}$$

We assume that the reward $r_h \geq 0$ for all h and that $\sum_{h=1}^{H} r_h \leq 1$ almost surely. We also assume that the Bellman matrix admits the following factorization. For all $h \in [H]$, there exist $\xi_h, \nu_h : \mathcal{F} \to \mathbb{R}^M$ such that $\mathcal{E}(f, \pi_{f'}, h) = \nu_h(f')^T \xi_h(f)$. Moreover, $\|\nu_h(f')\|_2 \leq \Psi, \|\xi_h(f)\|_2 \leq \Phi$ for all h, f, f' and denote $\zeta = \Psi\Phi$.

1.2 OLIVE

Algorithm 1 OLIVE

1: Estimate the predicted value $\hat{V}_f = \frac{1}{n_{est}} \sum_{i=1}^{n_{est}} \max_{a\in\mathcal{A}} f(x_1^i, a)$, where $x_1^i, i = 1, \ldots, n_{est}$ are n_{est} copies of the initial context
2: Let $\mathcal{F}_0 = \mathcal{F}$
3: **for** t=1,2,... **do**
4: Pick $f_t = \text{argmax}_{f\in\mathcal{F}_{t-1}} \hat{V}_f$ and $\pi_t = \pi_{f_t}$
5: Estimate the Bellman error $\sum_{h=1}^{H} \tilde{\mathcal{E}}(f_t, \pi_t, h)$ by sampling n_{eval} episodes $\{x_1^i, a_1^i, r_1^i, \ldots, x_H^i, a_H^i, r_H^i\}, i = 1, \ldots, n_{eval}$ with policy π_t
6: **if** $\sum_{h=1}^{H} \tilde{\mathcal{E}}(f_t, \pi_t, h) \leq \epsilon$ **then**
7: Terminate and return π_t
8: **end if**
9: Pick $h_t \in [H]$ such that $\tilde{\mathcal{E}}(f_t, \pi_t, h_t) \geq \frac{\epsilon}{H}$
10: Estimate the row associated with the roll-in policy π_t of the Bellman error matrix using importance sampling, i.e. collect n episodes $\{x_1^i, a_1^i, r_1^i, \ldots, x_H^i, a_H^i, r_H^i\}, i = 1, \ldots, n$ with $a_h^i = \pi_t(x_h^i)$ for $h \neq h_t$ and $a_{h_t}^i$ uniformly drawn from $\mathcal{A}$, and compute for each $f \in \mathcal{F}_{t-1}$

$$\hat{\mathcal{E}}(f, \pi_t, h_t) = \frac{1}{n} \sum_{i=1}^{n} \frac{I(a_{h_t}^i = \pi_f(x_{h_t}^i))}{1/K} \left(f(x_{h_t}^i, a_{h_t}^i) - r_{h_t}^i - f(x_{h_t+1}^i, \pi_f(x_{h_t+1}^i))\right)$$

11: Update $\mathcal{F}_t = \{f \in \mathcal{F}_{t-1} : |\hat{\mathcal{E}}(f, \pi_t, h_t)| \leq \phi\}$
12: **end for**

2 Theoretical Guarantee

The detail of the following analysis can be found in [JKA+16].

2.1 ϵ-suboptimality of the output

Lemma 1. *Let $V_f = \mathbb{E}_{x_1}[\max_{a\in\mathcal{A}} f(x_1, a)]$. Then $V_f - V^{\pi_f} = \sum_{h=1}^{H} \mathcal{E}(f, \pi_f, h)$.*

Proof:

$$
\begin{aligned}
&\sum_{h=1}^{H} \mathcal{E}(f, \pi_f, h) \\
=&\sum_{h=1}^{H} \mathbb{E}[f(x_h, a_h) - r_h - f(x_{h+1}, a_{h+1}) | a_{1:h+1} \sim \pi_f] \\
=&\sum_{h=1}^{H} \mathbb{E}[f(x_h, a_h) - r_h - f(x_{h+1}, a_{h+1}) | a_{1:H} \sim \pi_f] \\
=&\mathbb{E}[\sum_{h=1}^{H} (f(x_h, a_h) - r_h - f(x_{h+1}, a_{h+1})) | a_{1:H} \sim \pi_f] \\
=&\mathbb{E}[f(x_1, a_1) - \sum_{h=1}^{H} r_h | a_{1:H} \sim \pi_f] \\
=&V_f - V^{\pi_f}
\end{aligned}
$$

where the second last equality comes from the default $f(x_{H+1}, a) \equiv 0$. □

Proposition 2. *Let $f = \text{argmax}_{f\in\mathcal{F}_{t-1}} V_f$. If $V_f - V^{\pi_f} \le \epsilon$ and $f^* \in \mathcal{F}_{t-1}$, then $V^{\pi_f} \ge V^*_{\mathcal{F}} - \epsilon$.*

Proof: By Lemma 1

$$
\begin{aligned}
V^{\pi_f} =& V_f - \sum_{h=1}^{H} \mathcal{E}(f, \pi_f, h) \\
\ge& V_f - \epsilon \\
\ge& V_{f^*} - \epsilon = V^*_{\mathcal{F}} - \epsilon
\end{aligned}
$$

where the last equality is because f^* has zero Bellman error. □

2.2 Bound the Number of Epochs

Lemma 3. *Let $V \subset$ be a closed and bounded subset of $\mathbb{R}^d$ and $p \in \mathbf{R}^d$. Let B be any ellipsoid that is centered at the origin and encloses V. Suppose $\exists \nu \in V$ such that $|p^T\nu| \ge \kappa > 0$ and B^+ is the minimum volume ellipsoid that encloses $\{\nu \in B : |p^T\nu| \le \gamma\}$. If $\frac{\gamma}{\kappa} \le \frac{1}{3\sqrt{d}}$, then*

$$
\frac{vol(B^+)}{vol(B)} \le \frac{3}{5}.
$$

Lemma 4. *(The key lemma) Suppose $|\hat{\mathcal{E}}(f,\pi_t,h_t) - \mathcal{E}(f,\pi_t,h_t)| \le \phi$ throughout the algorithm, which implies that $f^* \in \mathcal{F}_t$ for all t. Further assume that for all $h \in [H]$, if whenever $h_t = h$ we have*

$$|\mathcal{E}(f_t,\pi_t,h_t)| \ge 6\sqrt{M}\phi := \frac{\epsilon}{H}.$$

Then the number of epochs such that $h_t = h$ is bounded by $M \log \frac{\zeta}{2\phi} / \log \frac{5}{3}$.

Proof: First let's define the notations. Let $I_1, I_2, \ldots, I_T$ be the epoch such that $h_t = h$. Define $I_0 = 0$. Let $p_\tau = \nu_h(f_{I_\tau})$ for $\tau = 1, 2, \ldots, T$. Let $U(\mathcal{F}_{T_\tau}) = \{\xi_h(f) : f \in \mathcal{F}_{T_\tau}\}$ for $\tau = 0, 1, \ldots, T$. Let $V_0 = \{\nu : \|\nu\|_2 \le \Phi\}$, and $V_\tau = \{\nu \in V_{\tau-1} : |p_\tau^T \nu| \le 2\phi\}$. Accordingly, let $B_\tau, \tau = 0, 1, \ldots, T$ be the minimum volume ellipsoid that encloses V_τ. Note that since V_τ is centered at the origin, so is B_τ.

Second we show that the volume of B_τ shrinks exponentially. Note that due to the first condition of this lemma and the criterion 11 of the algorithm, $U(\mathcal{F}_{T_\tau}) \subset V_\tau$. We apply Lemma 3 with $p = p_\tau, \gamma = 2\phi, \kappa = 6\sqrt{M}\phi, V = V_{\tau-1}, B = B_{\tau-1}$. Note that the criterion 9 implies that $p_\tau^T \xi_h(f_{I_\tau}) \ge \frac{\epsilon}{H} = 6\sqrt{M}\phi = \kappa$, where $\xi_h(f_\tau) \in U(\mathcal{F}_{T_{\tau-1}}) \subset V_{\tau-1}$. And it is clear that $\frac{\gamma}{\kappa} = \frac{2\phi}{6\sqrt{M}\phi} = \frac{1}{3\sqrt{M}}$, thus Lemma 3 is applicable and we have

$$\frac{vol(B_{\tau-1}^+)}{vol(B_{\tau-1})} \le \frac{3}{5}.$$

Compared with $B_{\tau-1}^+$, B_τ is the minimum volume enclosing ellipsoid of a smaller region, thus

$$\frac{vol(B_\tau)}{vol(B_{\tau-1})} \le \frac{vol(B_{\tau-1}^+)}{vol(B_{\tau-1})} \le \frac{3}{5}.$$

Since it is assumed that $\|\nu_h(f')\|_2 \le \Psi$ for all $f' \in \mathcal{F}$, it is easy to check that the ball of radius $\frac{2\phi}{\Psi}$ centered at the origin is always within B_τ and it is trivial that $B_0 = V_0$. This gives an upper bound of T, the number of epochs such that $h_t = h$

$$\left(\frac{3}{5}\right)^T \Phi^M \ge \left(\frac{2\phi}{\Psi}\right)^M,$$

that is

$$T \le M \log \frac{\zeta}{2\phi} / \log \frac{5}{3}.$$

This completes the proof. □

2.3 Sampling Complexity

Theorem 5. *To find a policy $\hat{\pi}$ such that $V^{\hat{\pi}} \ge V_{\mathcal{F}}^* - \epsilon$ with probability at least $1 - \delta$, the number of episodes of data required by OLIVE algorithm is*

$$O\left(\frac{M^2 H^3 K}{\epsilon^2} \log \frac{N}{\delta}\right),$$

Sketch of Proof: This sketch of proof is not rigorous in the sense that the constants that appear with $N, K, H, M, \delta, \epsilon$ are not carefully tuned. See [JKA+16] for detailed proof. The proof consists of first using Lemma 4 to split the total failure probability δ into step 1,5,10 of the algorithm, computing the number of samples required to succeed in each of these steps, and finally summing up to get the total sampling requirement. The number of required samples in each step can be established using standard concentration inequalities.

1. In step 1, to estimate V_f up to error ϵ for all f, it is enough to set $n_{est} = O\left(\frac{1}{\epsilon^2}\log\frac{N}{\delta}\right)$

2. In step 5, to estimate $\mathcal{E}(f_t, \pi_t, h)$ up to error $\frac{\epsilon}{H}$ for all h, it is enough to set $n_{eval} = O\left(\frac{H^2}{\epsilon^2}\log\frac{H}{\delta}\right)$

3. In step 10, to estimate $\mathcal{E}(f, \pi_t, h_t)$ up to error $\phi = \frac{\epsilon}{6H\sqrt{M}}$ for all f, it is enough to set $n = O\left(\frac{K}{\phi^2}\log\frac{N}{\delta}\right) = O\left(\frac{MH^2K}{\epsilon^2}\log\frac{N}{\delta}\right)$

Now due to Lemma 4 the algorithm terminates in $O(HM)$ epochs, thus the total sampling complexity is $O\left(\frac{M^2H^3K}{\epsilon^2}\log\frac{N}{\delta}\right)$. □

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.