Tiger Problem

March 5, 2014

State space (tiger lies left, tiger lies right)

$$\{S\} = \{s_0, s_1\} \tag{1}$$

Belief state

$$b = (x,\, 1-x) \tag{2}$$

Action (listen left; open left, open right)

$$\{a\} = \{a_0; a_1, a_2\} \tag{3}$$

Observation (tiger noise, no tiger noise)

$$\{z\} = \{z_0, z_1\} \tag{4}$$

Belief state transition (z observed)

$$b'(s')\Big|_{a,b,z} = \frac{1}{p(z|b,a)}\mathbb{P}[z|s',a]\sum_s \mathbb{P}[s'|s,a]b(s) \equiv b^{a,z} \tag{5}$$

Belief state transition (z not observed)

$$\mathbb{P}[b'\big|b,a] = \begin{cases} \sum_z \mathbb{P}[z\big|b,a] = \sum_z\sum_s \mathbb{P}[z\big|s,a]b(s)\ , & \text{for } \forall z \text{ s.t. } b' = b^{a,z} \\ 0\ , \quad \text{if no } z \text{ satisfies } b' = b^{a,z} \end{cases} \tag{6}$$

Value iteration (k indicates value iteration number)

$$V^{k+1}(b) = \max_a \left\{ \sum_s r(a,s)b(s) + \gamma \sum_z \max_{\mathbf{j}} \left(\sum_{s'} \alpha_{\mathbf{j}}^{k-1}(s')\mathbb{P}\left[z\big|s',a\right] \sum_s \mathbb{P}\left[s'\big|s,a\right] b(s) \right) \right\} \tag{7}$$

α-vector pool: $\Gamma^k \equiv \left\{(\alpha_1, a_{l_1})^k, \cdots, (\alpha_{\mathbf{j}}, a_{l_{\mathbf{j}}})^k, \cdots\right\}$
Given b, α-vector backup:

1. For $\forall a$, $\forall z$, find

$$\alpha^{k-1,*}\big|_{a,z} = \arg\max_{\alpha_{\mathbf{j}}^{k-1}\in\Gamma^{k-1}} \sum_{s'} \alpha_{\mathbf{j}}^{k-1}(s')\mathbb{P}\left[z\big|s',a\right] \sum_s \mathbb{P}\left[s'\big|s,a\right] b(s) \tag{8}$$

2. For $\forall a$, $\forall s$, find

$$\alpha^k(s)\big|_a = r(a,s) + \gamma \sum_z \sum_{s'} \alpha^{k-1,*}(s')\big|_{a,z}\mathbb{P}\left[z\big|s',a\right]\mathbb{P}\left[s'\big|s,a\right] \tag{9}$$

3. Find

$$a^* = \arg\max_a \sum_s \alpha^k(s)\big|_a b(s) \tag{10}$$

4. **[optional]** Append α-vector pool (with redundancy)

$$\bar{\Gamma}^k \leftarrow \left\{\Gamma^{k-1}, (\alpha^k\big|_{a^*}, a^*)\right\} \tag{11}$$

Definition 1. *Dominate subset of vectors:*
Let $\bar{\Gamma} = \{\alpha_1, \cdots, \alpha_q\}$ be a set of vectors in $\mathbb{R}^n$, b be a vector in $\mathbb{R}^n$. Γ is called the dominate subset of $\bar{\Gamma}$, iff for $\forall \alpha_i \in \Gamma \subseteq \bar{\Gamma}$, $\exists b$, s.t.

$$\alpha_i^T b > \alpha_{\{1,\cdots,q\}\setminus i}^T b \tag{12}$$

subject to

$$\begin{aligned} I_{n\times n} b \geq 0 \\ \begin{pmatrix}1, & \cdots, & 1\end{pmatrix} b = 1 \end{aligned} \tag{13}$$

Γ can be obtained from $\bar{\Gamma}$ by solving a linear programming problem by e.g. simplex method.
To implement point-based value iteration on all $\mathcal{R}^L(b_0)$ ($\mathcal{R}$ indicates reachable belief states, L indicates maximum depth)

1. **Initialize** Belief tree $T(b_0)$ with maximum depth L, consisted of *action nodes* and *belief nodes.*
2. **repeat**
3. Choose all belief nodes b_l from $T(b_0)$
4. For each b_l, backup $\bar{\Gamma}$
5. $\Gamma \leftarrow \bar{\Gamma}$
6. **until** the max increament of $V(b_l)$ is smaller than ϵ
7. **return** Γ

Instead of the above approach, consider Perseus algorithm

Perseus Algorithm

1. **initialize**
2. Initialize Belief tree $T(b_0)$ with maximum depth L
3. $B \leftarrow$ random subset of non-leaf belief nodes of $T(b_0)$
4. Lower bound of α-vectors
$$\alpha \leftarrow \frac{\min\{r\}}{1-\gamma}(1,1) \tag{14}$$
5. Iteration count $k \leftarrow 0$
6. α-vector pool $\Gamma^0 \leftarrow \{\alpha\}$
7. **repeat**
8. **for** b **in** B **:**
9. Store $\alpha^k(b) \leftarrow \arg\max_{\alpha\in\Gamma^k} \langle\alpha, b\rangle$
10. Store $V^k(b) \leftarrow \left\langle\alpha^k(b), b\right\rangle$
11. $\Gamma^{k+1} \leftarrow \varnothing$
12. Not-improved belief points $\tilde{B} \leftarrow B$

while $\tilde{B}$ **is not** $\varnothing$ **:**

  $b \leftarrow$ a random sample from $\tilde{B}$

  $\alpha \leftarrow \text{backup}(b, \Gamma^k)$

  if $\langle \alpha, b \rangle < V^k(b)$ **:**

    $\alpha \leftarrow \alpha^k(b)$

  for b' **in** $\tilde{B}$ **:**

    if $\langle \alpha, b' \rangle \geq V^k(b')$ **:**

      $\tilde{B} \leftarrow \tilde{B} \setminus b'$

  $\Gamma^{k+1} \leftarrow \Gamma^{k+1} \cup \alpha$

$k \leftarrow k+1$

until convergence

1 Continuous poMDP

Bellman equation

$$V(b) = \sup_{a \in A} \left\{ \langle r_a, b \rangle + \gamma \int_z \mathbb{P}[z \big| b, a] V\left(b^{a,z}\right) \mathrm{d}z \right\} \tag{15}$$