EVB Decomposition

Change in the value function due to learning after taking action a^*:

$$\begin{aligned} v(ba^*) - v(b) =& \sum_{b'} p(b' \mid b, a^*)\Big(v(b') - v(b)\Big) \\ =& \sum_{b'} p(b' \mid b, a^*)\Big(\sum_a \pi(a \mid b')q(b', a) - \sum_a \pi(a \mid b)q(b, a)\Big) \\ =& \sum_{b'} p(b' \mid b, a^*)\sum_a \Big(\big(\pi(a \mid b') - \pi(a \mid b)\big)q(b', a) \\ &+\pi(a \mid b)\big(q(b', a) - q(b, a)\big)\Big) \end{aligned} \tag{1}$$

Expanding $q(b', a) - q(b, a)$:

$$\begin{aligned} q(b', a) - q(b, a) =& \sum_{b''} p(b'' \mid b', a)\big[r(b', a) + \gamma v(b'')\big] \\ &- \sum_{b'} p(g' \mid b, a)\big[r(b, a) + \gamma v(g')\big] \\ =& r(b', a) + \gamma \sum_{b''} p(b'' \mid b', a)v(b'') \\ &-r(b, a) + \gamma \sum_{g'} p(g' \mid b, a)v(g') \\ =& \quad \underbrace{r(b', a) - r(b, a)}_{\substack{\text{Difference in the expected} \\ \text{immediate return}}} \quad \underbrace{+\gamma\Big[\sum_{b''} p(b'' \mid b', a)v(b'') - \sum_{g'} p(g' \mid b, a)v(g')\Big]}_{\substack{\text{Difference in the expected} \\ \text{future return}}} \end{aligned} \tag{2}$$

So overall the EVB decomposes as:

$$\begin{aligned} v(ba^*) - v(b) =& \mathbb{E}_{b' \sim p(b' \mid b, a^*)}\Big[\sum_a \big(\pi(a \mid b') - \pi(a \mid b)\big)q(b', a) \\ &+\mathbb{E}_{a \sim \pi(a \mid b)}\big[r(b', a) - r(b, a)\big] \\ &+\mathbb{E}_{a \sim \pi(a \mid b)}\big[\gamma \sum_{b''} p(b'' \mid b', a)v(b'') - \gamma \sum_{g'} p(g' \mid b, a)v(g')\big]\Big] \end{aligned} \tag{3}$$

One last thing to consider is how the prioritisation of distal experiences should differ from those that are more immediate. This also has implications for how likely those experiences are to occur according to the current model.

For instance, if the agent considers updating $v(b)$ towards the value that would result from taking a particular action from that belief state – say, $v(ba^*)$ – the EVB associated with that update needs to be weighted by the probability of transitioning into belief state b in the first place (i.e., from the current root of the tree).

$$\begin{aligned} v(ba^*) - v(b) = p(b_{\text{root}} \to b) \times \Big(& \mathbb{E}_{b' \sim p(b' \mid b, a^*)}\Big[\sum_a \big(\pi(a \mid b') - \pi(a \mid b)\big)q(b', a) \\ &+\mathbb{E}_{a \sim \pi(a \mid b)}\big[r(b', a) - r(b, a)\big] \\ &+\mathbb{E}_{a \sim \pi(a \mid b)}\big[\gamma \sum_{b''} p(b'' \mid b', a)v(b'') - \gamma \sum_{g'} p(g' \mid b, a)v(g')\big]\Big]\Big) \end{aligned} \tag{4}$$

Simulations

Prioritisation pattern with horizon $h = 2$.