Assignment 5 - GVFs and Successor Representation

Jianing Sun 260791202

- Common Question

Question: Explain how SR can be expressed in the GVF framework (also see corresponding section in S&B if necessary). Explain this connection using the GVF terminology : cumulant (this is not the notion of "moments" of a prob. distribution), termination condition, etc.

Answer: In General Value Functions (GVF), whatever signal is added up in a value-function-like prediction, we call it the *cumulant* of that prediction. Denote $C_t \in \mathbb{R}$ as the **cumulant** signal, and **terminal function** $\gamma : \mathcal{S} \mapsto [0, 1]$. Then by definition, a **general value function**, or GVF, is written

$$v_{\pi,\gamma,C}(s) = \mathbb{E}\Big[\sum_{k=t}^{\infty} C_{k+1} \prod_{i=t+1}^{k} \gamma(S_i) \Big| S_t = s, A_{t:\infty} \sim \pi\Big]$$

Different choices of cumulant signal C lead to different characterizations of the Markov Chain. For example, if C specfies the reward received on entering state s, then $v_{\pi,\gamma,C}(s)$ correponds to the expected discounted future return, or value, the standard target of RL. In the context of successor representation (SR), **successor representation can be expressed as part of the GVF framework in this way:**

$$v_{\pi,\gamma,C}(s) = \sum_{s' \in S} \phi_\pi(s, s')C(s)$$

where ϕ_π is the successor representation (SR), which **encodes the expected discounted future visitations of each state s' along trajectories originating in state s.** Wherein C(s) is the cumulant.

- Trace 1

1. Starting from the definition of above, derive Bellman equations for phi*pi. Note : phi*\pi(s) has essentially the meaning of a "value function".

$$\phi_\pi(s) = e_s^T(I - \gamma P_\pi)^{-1} = e_s^T \sum_{t=0}^{\infty} \gamma^t P_\pi = \mathbb{E}\Big[e_s^T \sum_{t=0}^{\infty} \gamma^t\Big]$$

$$
\begin{aligned}
\Rightarrow \phi_\pi(s,s') &= \mathbb{E}\Big[\sum_{t=0}^{\infty}\gamma^t \mathbb{1}(S_t,s')\Big|S_o=s\Big] \\
&= \mathbb{E}\Big[\mathbb{1}(S_0,s')+\sum_{t=1}^{\infty}\gamma^t \mathbb{1}(S_t,s')\Big|S_0=s\Big] \\
&= \mathbb{E}\Big[\mathbb{1}(S_0,s')+\sum_{t=0}^{\infty}\gamma^{t+1} \mathbb{1}(S_{t+1},s')\Big|S_0=s\Big] \\
&= \mathbb{E}\Big[\mathbb{1}(S_0,s')+\gamma\sum_{t=1}^{\infty}\gamma^t \mathbb{1}(S_{t+1},s')\Big|S_0=s\Big] \\
&= \mathbb{E}\Big[\mathbb{1}(S_0,s')+\gamma\phi_\pi(S_1)\Big|S_0=s\Big]
\end{aligned}
$$

where $\mathbb{1}(s,s')$ is an indicator where all elements are zero except 1 for element s'. In other reference paper (https://papers.nips.cc/paper/5340-design-principles-of-the-hippocampal-cognitive-map.pdf), they usually denote successor representation (SR) with two parameters as M(s,s'). We do not need to marginalize over s' for next step derivation of the successor representation's Bellman Equation. We only need to marginalize over S_1 :

$$
\begin{aligned}
\phi_\pi(s) &= \sum_{s'} T_{s,s'}\left[\mathbb{1}(s,s')+\gamma\phi_\pi(s')\right] \\
&= \sum_{s'} P(s'|s,a)\sum_a \pi(a|s)\left[\mathbb{1}(s,s')+\gamma\phi_\pi(s')\right] \\
&= \sum_a \pi(a|s)\Big[\sum_{s'} P(s'\Big|s,a)\mathbb{1}(s,s')+\gamma\sum_{s'} P(s'\Big|s,a)\phi_\pi(s')\Big]
\end{aligned}
$$

2. Derive Bellman equations for state and action dependent phi*pi(s,a). Let’s call this psi*\pi(s,a) to avoid using the same notation. You can now think of psi*pi(s,a) as your “Q*\pi(s,a)” and where the “rewards” are vector-valued.

Answer:

assume π fixed,

$$
\begin{aligned}
Q^{\pi}(s,a) &= \mathbb{E}\Big[R(S_0,A_0)+\gamma R(S_1,A_1),+\ldots\Big|S_0=s,A_0=a,Ai\sim\pi(\cdot|S_i)\Big]\\
&= \mathbb{E}\Big[\phi(S_0,A_0)w+\gamma\phi(S_1,A_1)w+\ldots\Big|S_0=s,A_0=a,Ai\sim\pi(\cdot|S_i)\Big]\\
&= \psi^{\pi}(s,a)\cdot w\\
\Rightarrow \psi^{\pi}(s,a) &= \mathbb{E}_{\pi}\Big[\sum_{i=1}^{\infty}\gamma^{i-1}\phi(S_i,A_i)\Big|S_0=s,A_0=a\Big]
\end{aligned}
$$

Where w are parameters to learn.

Special case: for Tabular case, ϕ means indicator variables

$$
\Rightarrow \psi^{\pi}(s,a)=\phi(s,a)+\gamma\sum_{s'}P(s'|s,a)\sum_{a'}\pi(a'|s')\psi^{\pi}(s',a') \tag{1}
$$

Therefore, Eq.(1) is the bellman equation for $\psi^{\pi}(s,a)$ which is state-action dependent.

3. Show that the “Bellman operator” underlying the evaluation equations in 1. is indeed a contraction.

Linear Case:

$$
\phi_{\pi}(s,s')=\sum_{a}\pi(a|s)\Big[\sum_{s'}P(s'\Big|s,a)\mathbb{1}(s,s')+\gamma\sum_{s'}P(s'|s,a)\phi_{\pi}(s_1)\Big]
$$

$$
\forall s,\phi_{k+1}(s)\leftarrow\sum_{a}\pi(a|s)\Big(\sum_{s'}P(s'\Big|s,a)\mathbb{1}(s,s')+\gamma\sum_{s'}P(s'\Big|s,a)\phi_{k}(s')\Big)
$$

$$
\begin{aligned}
\phi_{k+1}(s)-\phi_{\pi}(s) &= \sum_{a}\pi(a|s)\sum_{s'}P(s'\Big|s,a)\mathbb{1}(s,s')+\sum_{a}\pi(s|a)\gamma\sum_{s'}P(s'\Big|s,a)\phi_{k}(s')\\
&-\sum_{a}\pi(a|s)\sum_{s'}P(s'\Big|s,a)\mathbb{1}(s,s')+\sum_{a}\pi(s|a)\gamma\sum_{s'}P(s'\Big|s,a)\phi_{\pi}(s')\\
&=\gamma\sum_{a}\pi(a|s)\sum_{s'}P(s'\Big|s,a)\Big[\phi_{k}(s')-\phi_{\pi}(s')\Big]\\
&\leq\gamma\sum_{a}\pi(a|s)\sum_{s'}P(s'\Big|s,a)\max_{s'}\Big[\phi_{k}(s')-\phi_{\pi}(s')\Big]
\end{aligned}
$$

denote $\varepsilon_{k+1}(s) = \phi_{k+1}(s) - \phi_\pi(s)$

$$\varepsilon_k(s) = \left\|\phi_k(s) - \phi_\pi(s)\right\|_{k\to\infty}$$

Hence,

$$\begin{aligned}
\phi_{k+1}(s) - \phi_\pi(s) &\le \gamma \sum_a \pi(a|s) \sum_{s'} P(s'\big|s,a) \max_{s'} \left[\phi_l(s') - \phi_\pi(s')\right] \\
&\le \gamma\varepsilon_k \sum_a \pi(a|s) \sum_{s'} P(s'\big|s,a) \\
&\le \gamma\varepsilon_k
\end{aligned}$$

That is $\varepsilon_{k+1} \le \gamma\epsilon_k$, and $\gamma \in (0,1)$

$$\Rightarrow \varepsilon_{k+1} \le \gamma^k \varepsilon_0 \xrightarrow{k\to\infty} 0$$

Therefore, it is indeed a contraction Bellman operator. Non-linear case is the same derivation process as the one in linear case, just under the maximization condition.