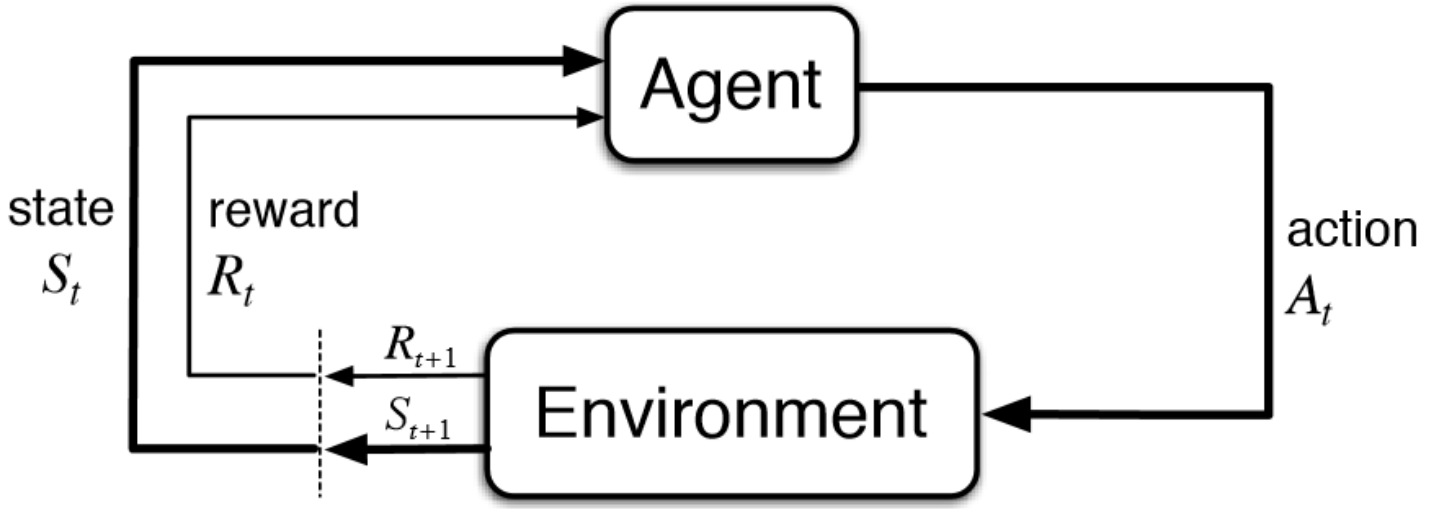
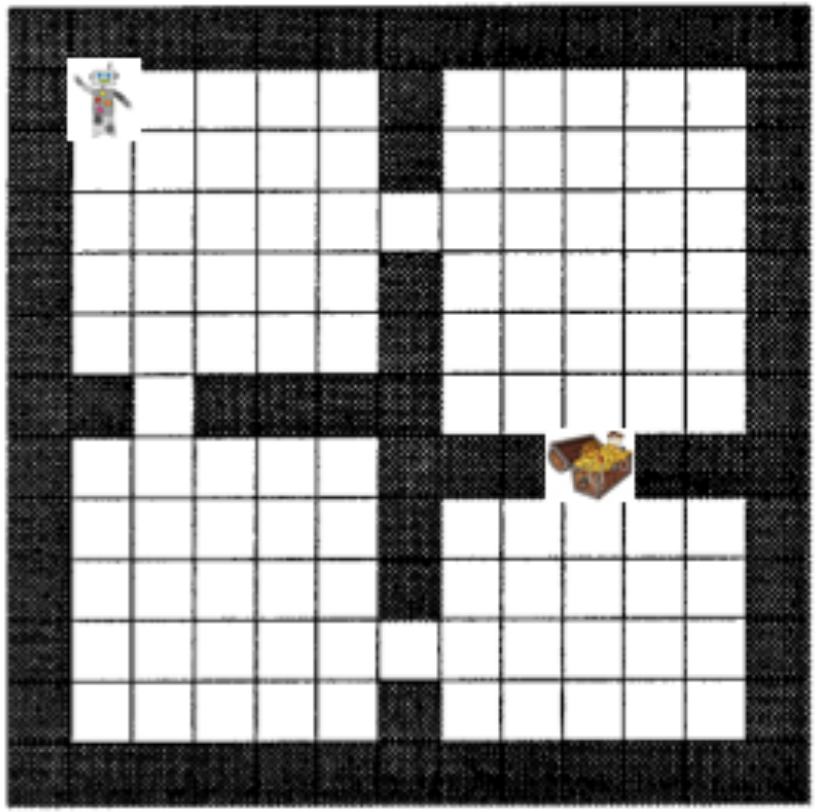
Playing with temporal abstraction for reinforcement learning

Author: Théophile Gervet

Advisor: Doina Precup

What is reinforcement learning?





$$\mathcal{P}^a_{ss'} = P(S_{t+1} = s' \mid S_t = s, A_t = a)$$

$$\pi(a \mid s) = P(A_t = a \mid S_t = s)$$

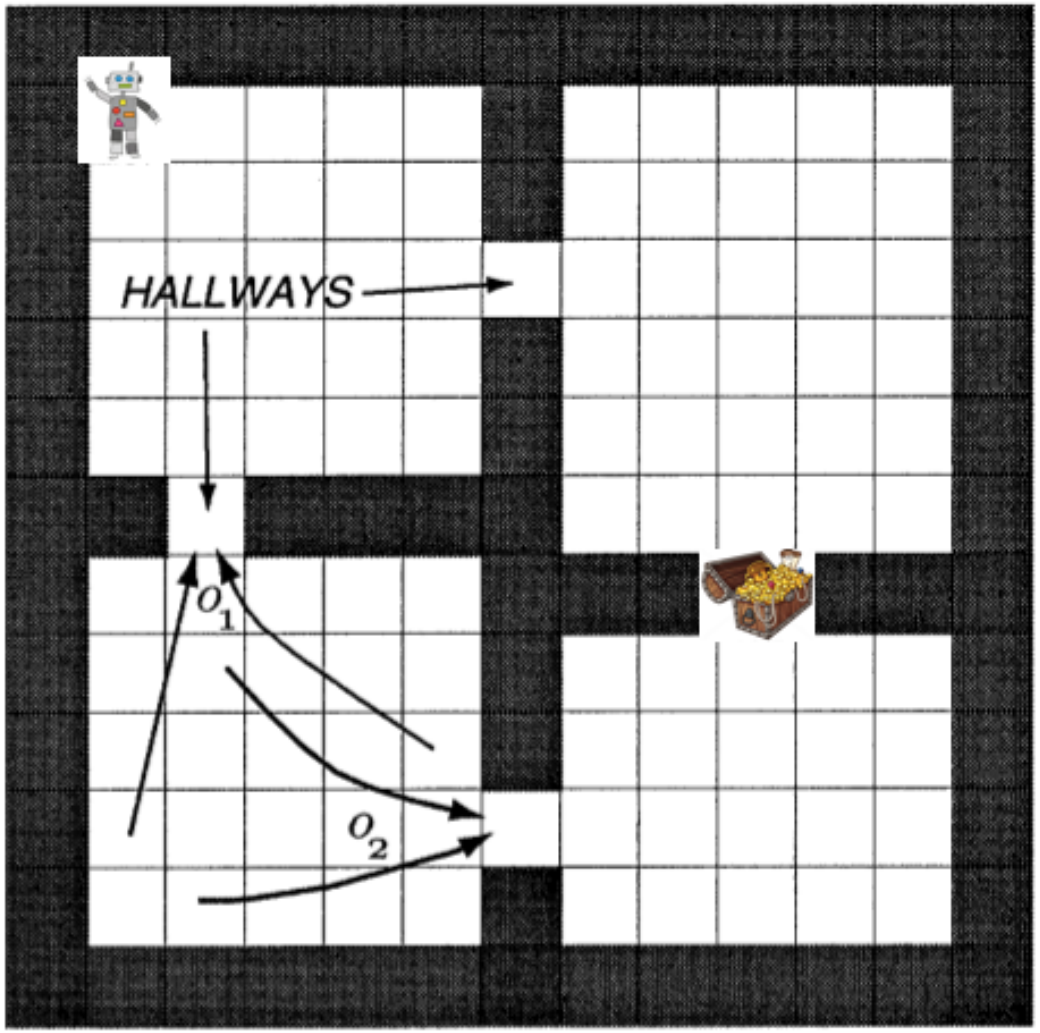
$$\mathcal{R}^a_s = E[R_{t+1} \mid S_t = s, A_t = a]$$

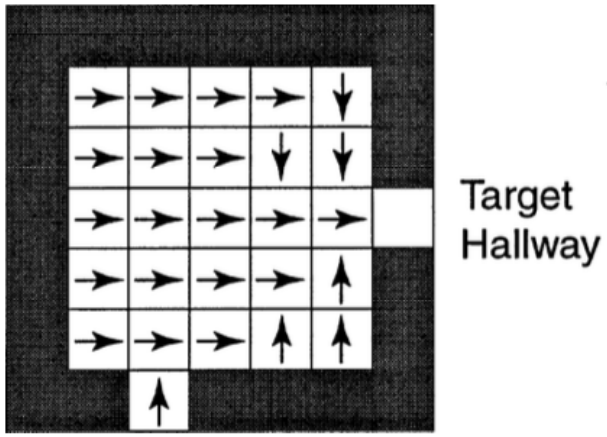
Bellman equations

$$
\begin{aligned}
V^{\pi}(s) &= E[R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} \ldots \mid S_t = s] \\
&= E[R_{t+1} + \gamma V^{\pi}(S_{t+1})] \\
&= \sum_a \pi(a|s)[\mathcal{R}_s^a + \gamma \sum_{s'} \mathcal{P}_{ss'}^a V^{\pi}(s')]
\end{aligned}
$$

$$
\begin{aligned}
V^{*}(s) &= \max_{\pi} V_{\pi}(s) \\
&= \max_{a} E[R_{t+1} + \gamma V^{*}(S_{t+1}) \mid S_t = s, A_t = a] \\
&= \max_{a} [\mathcal{R}_s^a + \gamma \sum_{s'} \mathcal{P}_{ss'}^a V^{*}(s')]
\end{aligned}
$$

What is temporal abstraction?



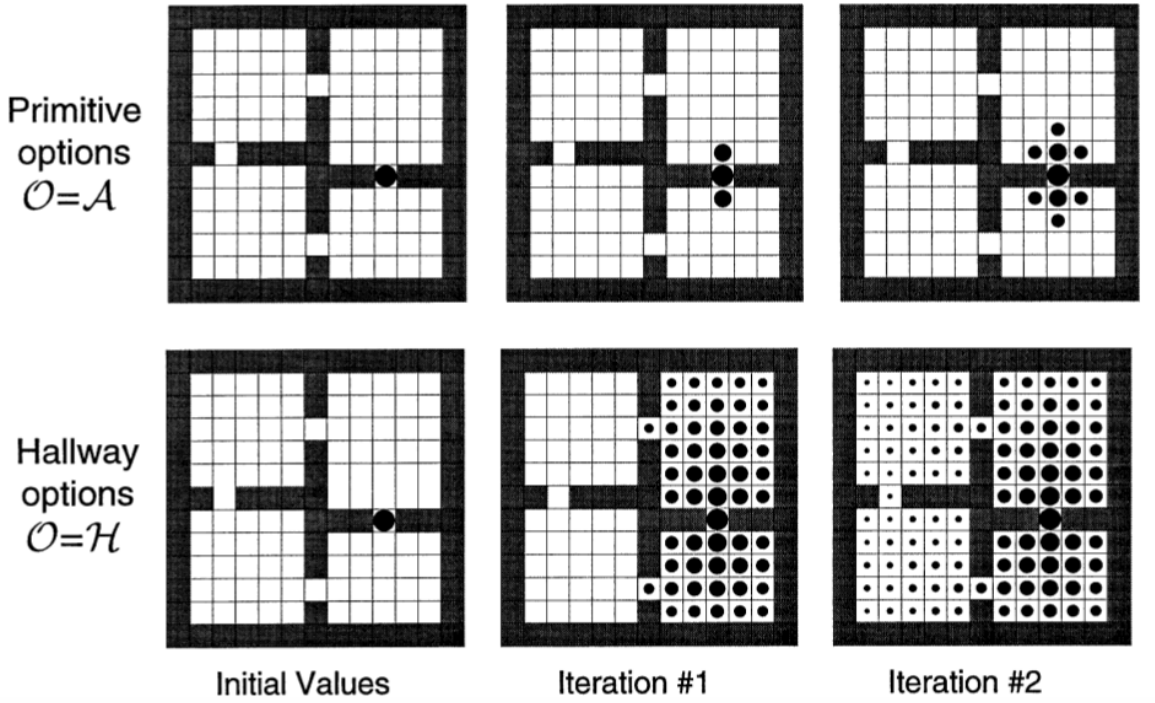


$$o = (\pi, \beta, I)$$

$$\begin{cases} \pi : S \times A \to [0, 1] \\ \beta : S \to [0, 1] \\ I \subseteq S \end{cases}$$

$$\mu(o \mid s) = P(O_t = o | S_t = s)$$

Planning with options



$$\mathcal{R}_s^o = E[R_{t+1}+\gamma R_{t+2}+\ldots+\gamma^{k-1}R_{t+k} \mid o \text{ initiated in state } s \text{ at time step } t]$$

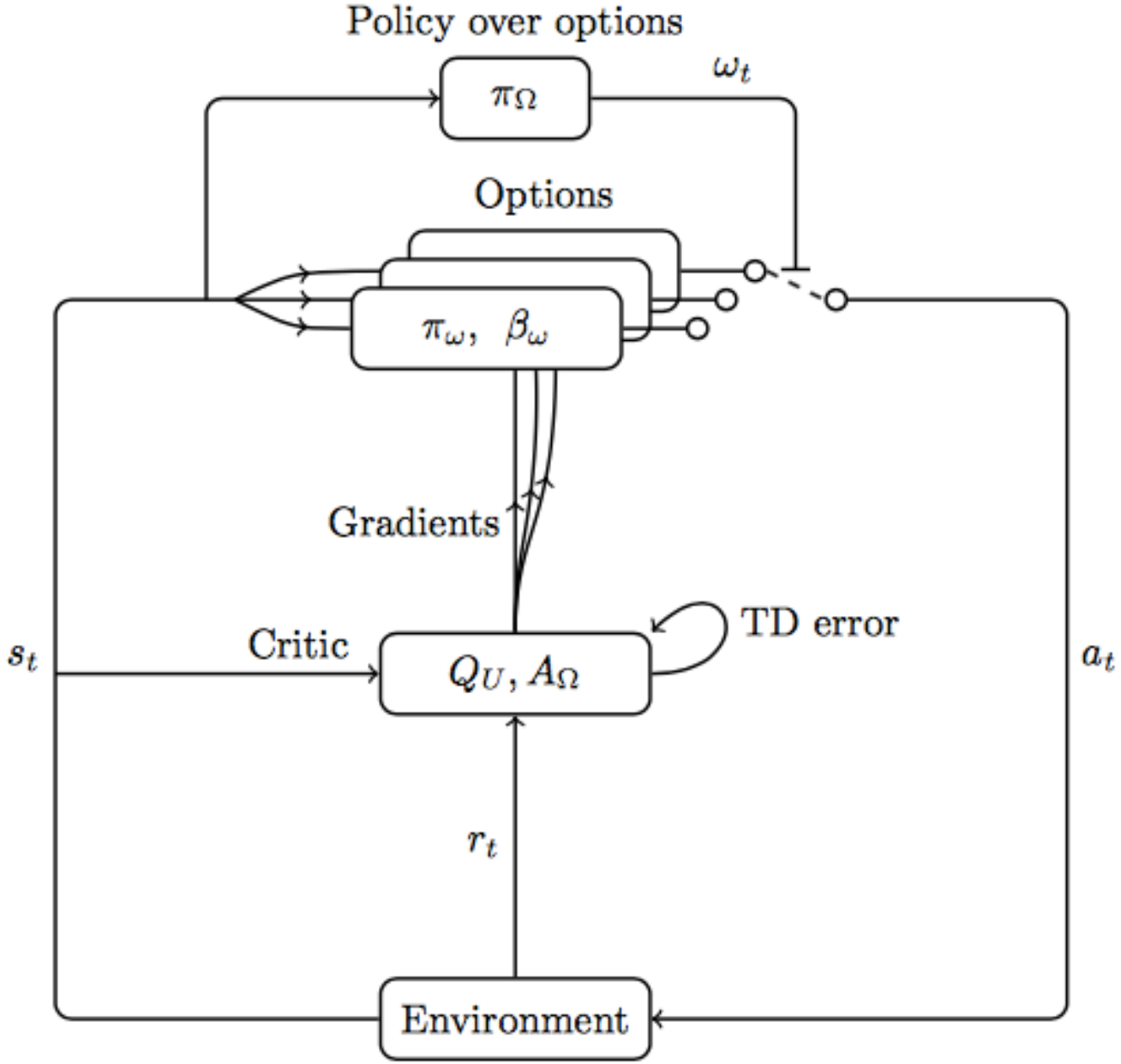
$$\mathcal{P}_{ss'}^o = \sum_{k=1}^{\infty} \gamma^k P\{o \text{ terminates in } s' \text{ after } k \text{ steps}\}$$

$$V^\mu(s) = E[R_{t+1} + \ldots + \gamma^{k-1}R_{r+k} + \gamma^k V^\mu(S_{t+k})]$$

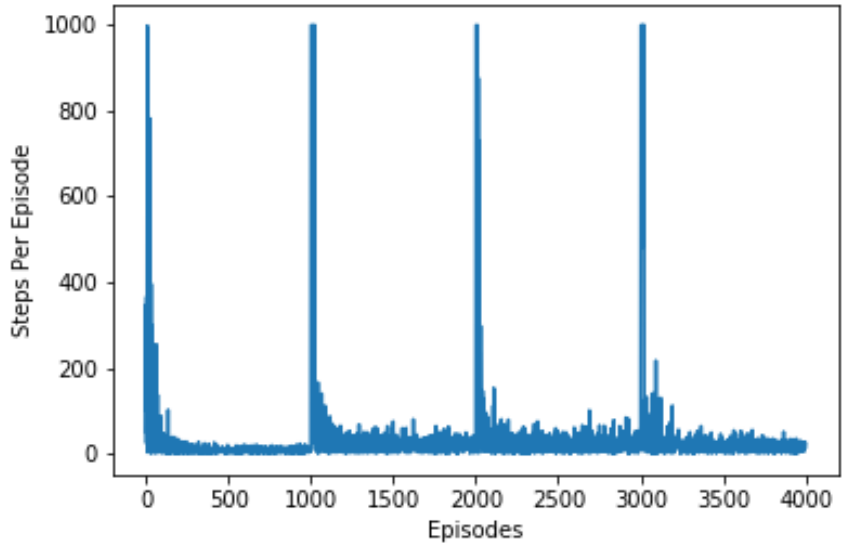
$$= \sum_o \mu(o \mid s)[\mathcal{R}_s^o + \sum_{s'} \mathcal{P}_{ss'}^o V^\mu(s')]$$

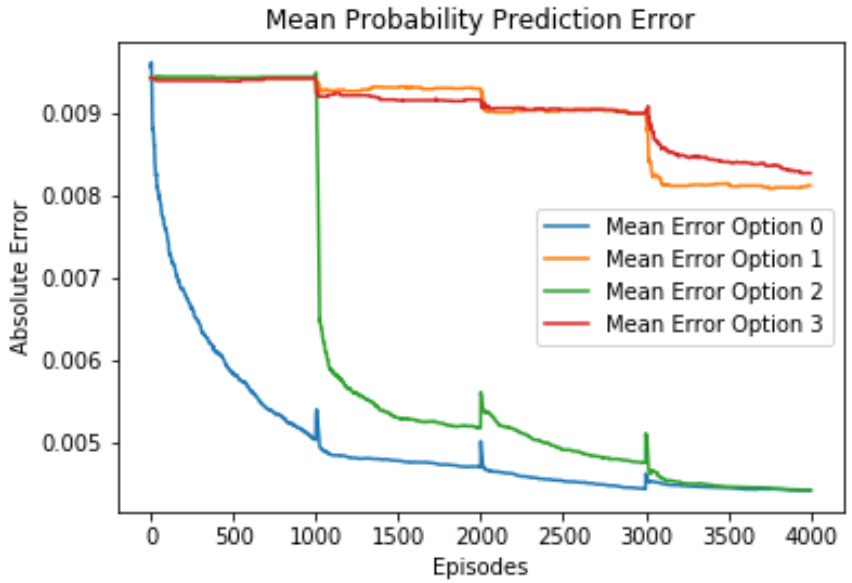
Option-critic

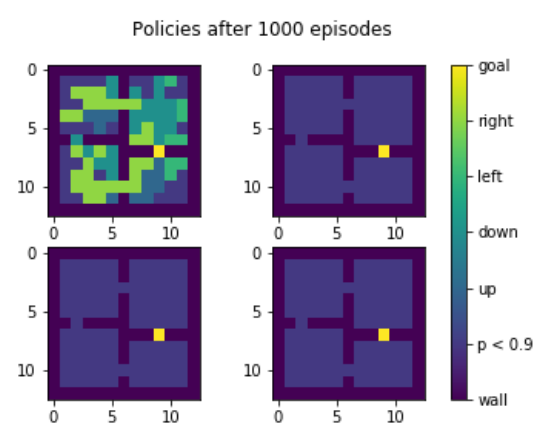
- Learn options directly from data
- Policy gradient algorithm
- Similar to Actor-Critic

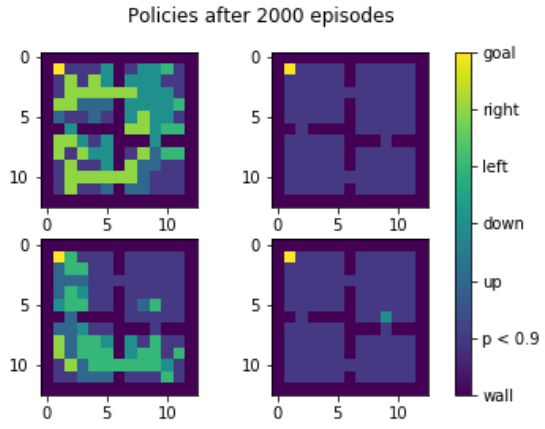


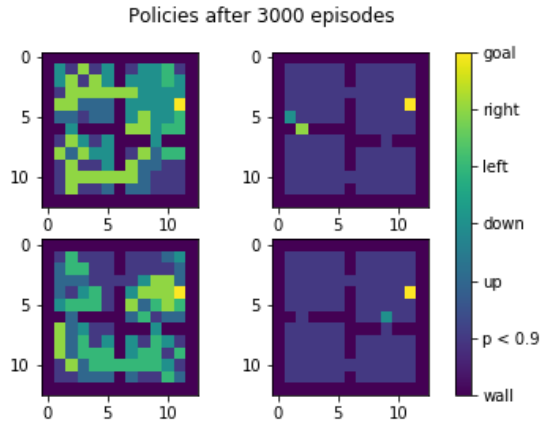
Transfer learning + model learning experiments

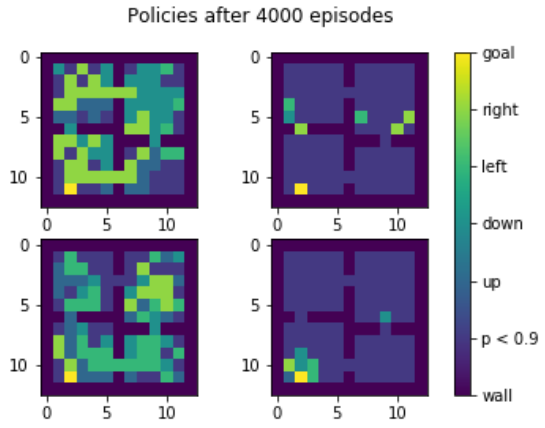












References

- *Between MDPs and Semi-MDPs: A framework for temporal abstraction in reinforcement learning*, Richard S. Sutton, Doina Precup, Satinder Singh – Artificial Intelligence, 1999

- *The Option-Critic Architecture*, Pierre-Luc Bacon, Jean Harb, Doina Precup - JMLR 2016