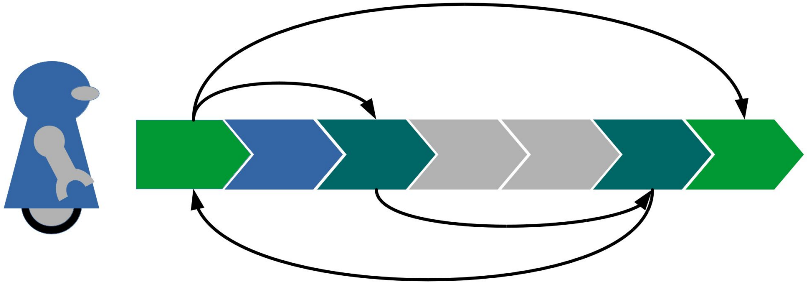
Successor Features for Transfer in Reinforcement Learning

André Barreto   Will Dabney   Rémi Munos   Jonathan Hunt   Tom Schaul   David Silver   Hado van Hasselt

Poster #9 at Pacific Ballroom

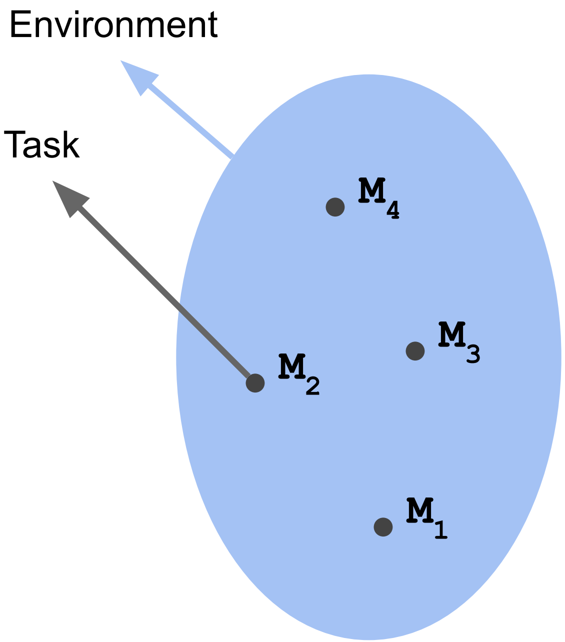
DeepMind

Transfer in Reinforcement Learning



- **Goal**: general framework for transfer in RL
- Fundamental design principles:
  - Exchange of information should take place whenever useful
  - Transfer should be seamlessly integrated within the RL process

Problem definition



- Environment is a **set** of MDPs
- Each MDP M_i is a **task**
- The only difference between the MDPs is the **reward function** r_i:

$$r_i(s, a, s') = \boldsymbol{\phi}(s, a, s')^\top \mathbf{w}_i$$

Proposed solution

Our solution builds on two ideas:

- Generalised policy improvement (GPI)
- Successor features (SFs)

Generalised policy improvement

For a fixed reward function:

$$\left.\begin{array}{l} \pi_1 \rightarrow Q^{\pi_1} \\ \pi_2 \rightarrow Q^{\pi_2} \\ \qquad \vdots \\ \pi_n \rightarrow Q^{\pi_n} \end{array}\right\} \boxed{\textbf{GPI}} \rightarrow \pi, \ \text{such that } Q^{\pi} \geq Q^{\pi_i}$$

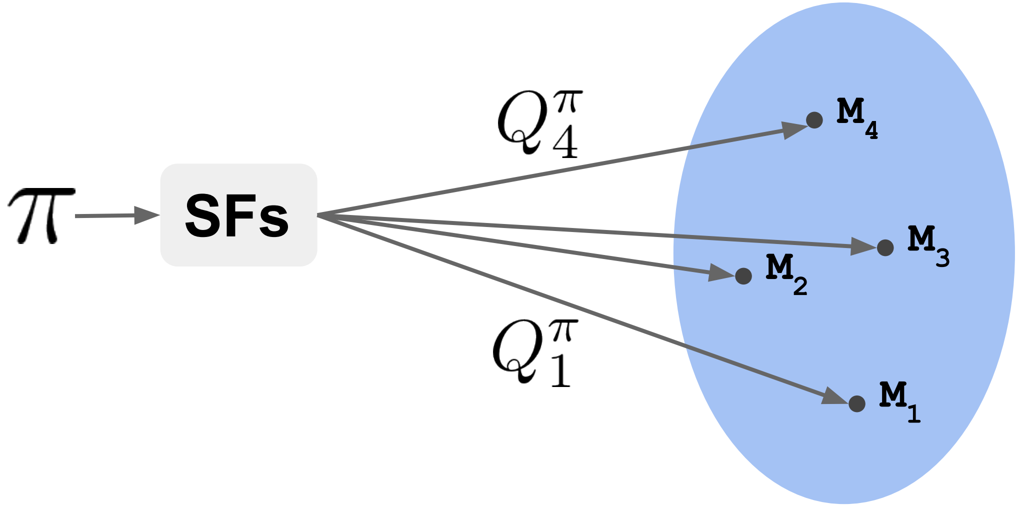
Generalised policy improvement

For a fixed reward function:

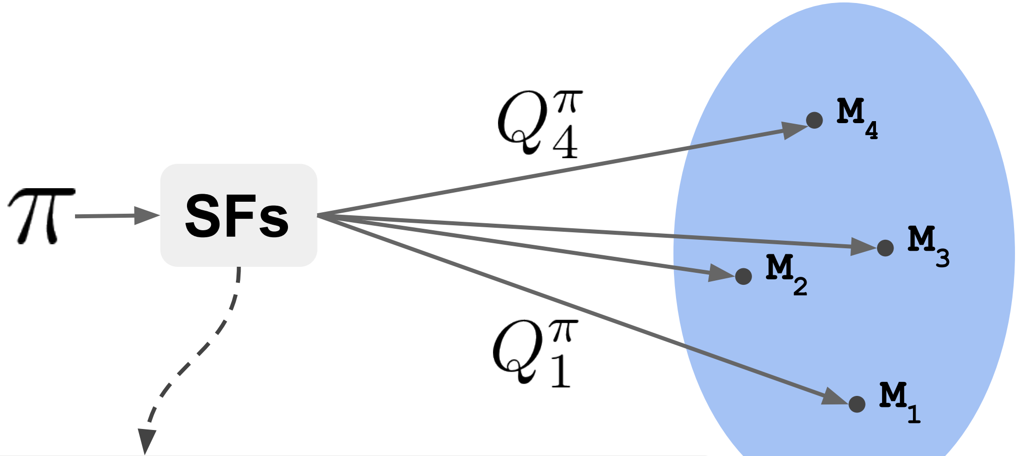
$$\left.\begin{array}{l} \pi_1 \rightarrow Q^{\pi_1} \\ \pi_2 \rightarrow Q^{\pi_2} \\ \quad\vdots \\ \pi_n \rightarrow Q^{\pi_n} \end{array}\right\} \textbf{GPI} \rightarrow \pi, \ \text{such that } Q^{\pi} \geq Q^{\pi_i}$$

$$\pi(s) \in \text{argmax}_a \text{max}_i Q^{\pi_i}(s, a)$$

Successor Features



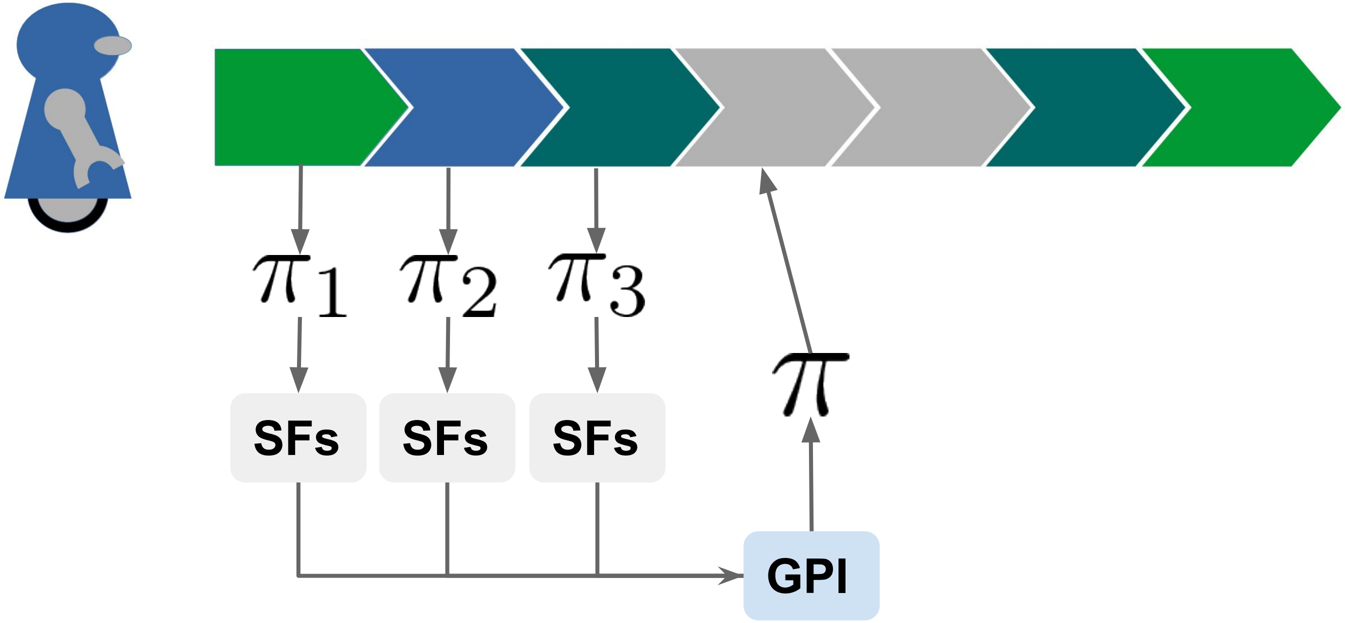
Successor Features



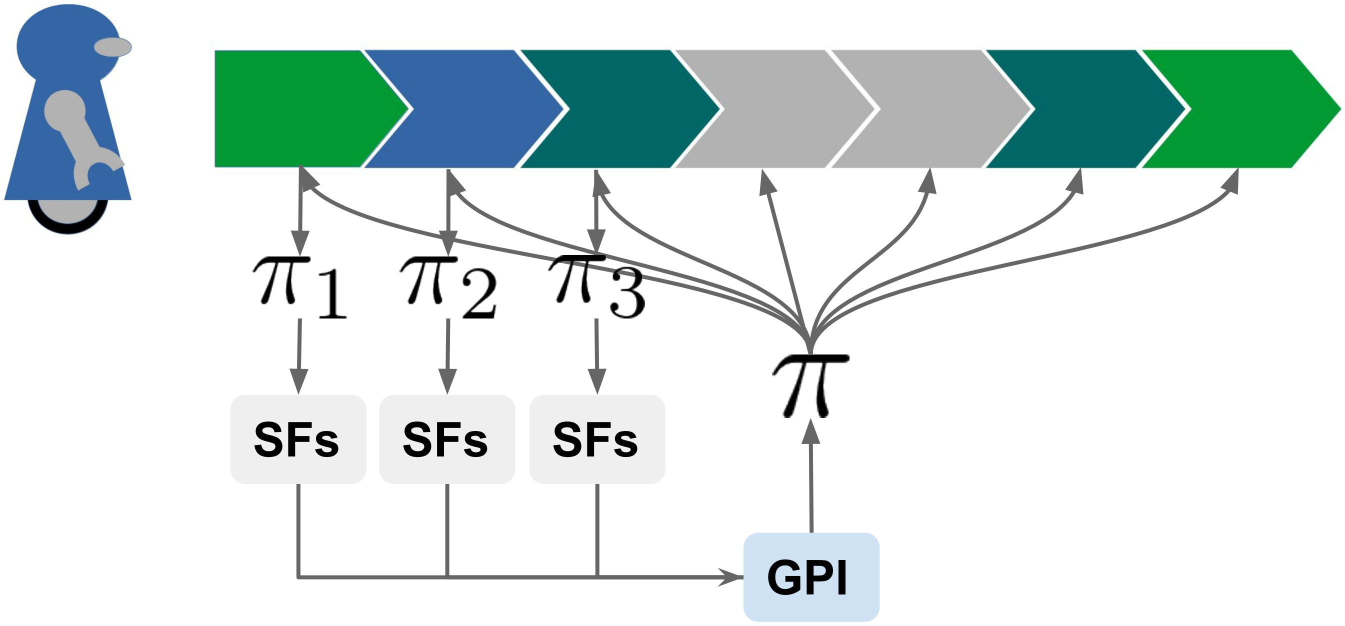
$$\psi^\pi(s,a) = \mathrm{E}\left[\phi(s,a,S') + \gamma\psi^\pi(S',\pi(S'))\right]$$

$$Q^\pi(s,a) = \psi^\pi(s,a)^\top \mathbf{w}$$

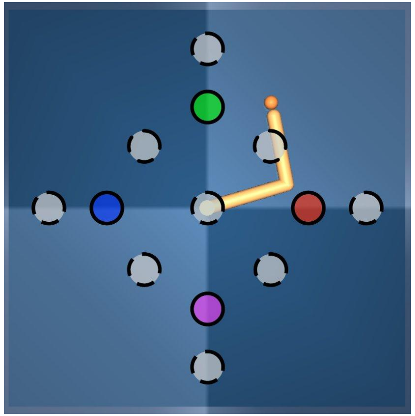
SFs + GPI

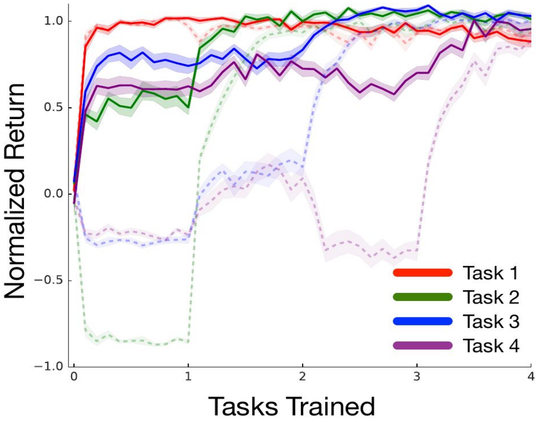


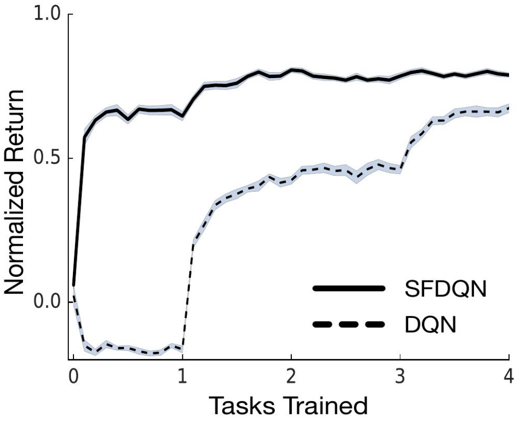
SFs + GPI



SFs + GPI







Generalisation across tasks (seen and novel)

Successor Features for Transfer in Reinforcement Learning

André Barreto Will Dabney Rémi Munos Jonathan Hunt Tom Schaul David Silver Hado van Hasselt

Poster #9 at Pacific Ballroom

DeepMind