Proto-Value Networks: Scaling Representation Learning with Auxiliary Tasks

Google Research
Mila

Jesse Farebrother, Joshua Greaves*, Rishabh Agarwal, Charline Le Lan, Ross Goroshin, Pablo Samuel Castro, Marc G. Bellemare*

Temporal Representations with Proto-Value Functions (PVFs)

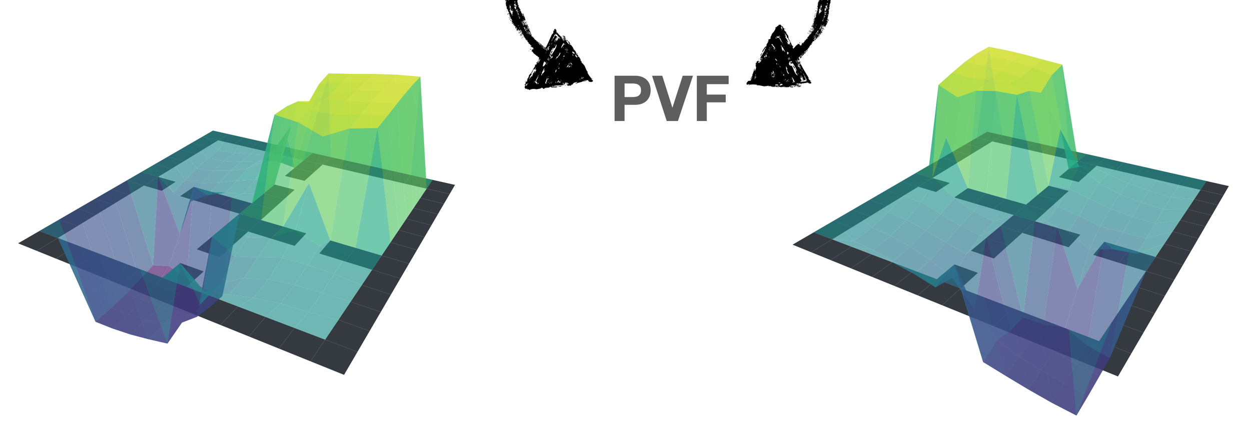
- PVFs can be thought of as capturing large-scale temporal properties of the environment.
- Encodes the structure of the MDP at different spatial scales.
- PVFs can't be easily scaled to large state-spaces.

$$L = D - A$$

Degree Matrix — Adjacency matrix

$$L\mathbf{v} = \lambda\mathbf{v}$$

PVF



$\mathbf{v}_1$ $\mathbf{v}_2$

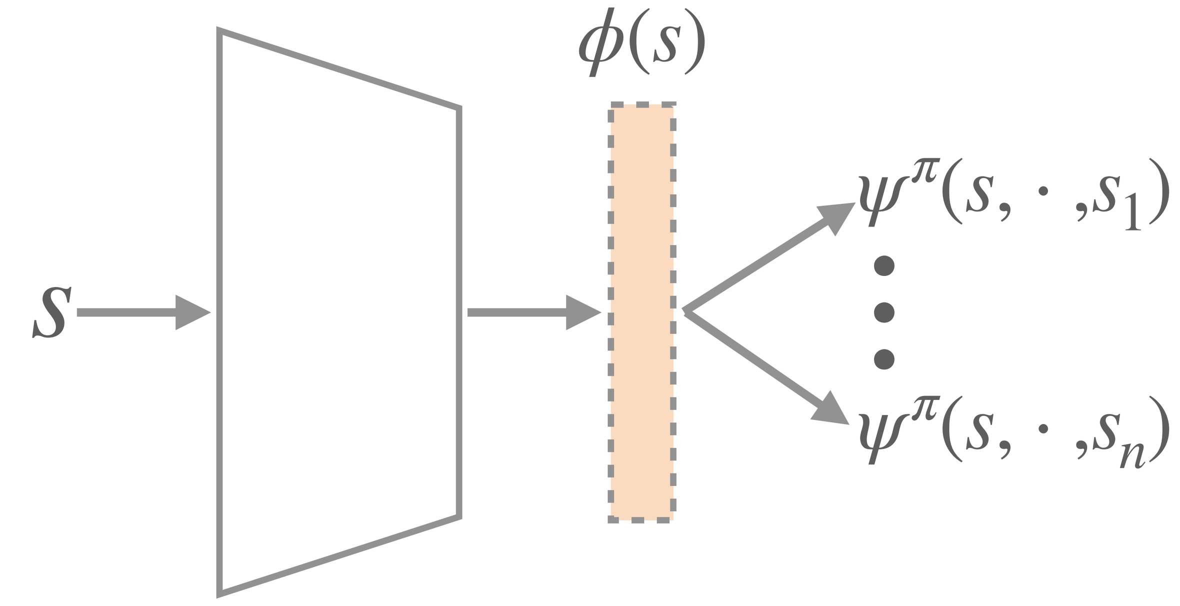
PVFs & The Successor Representation (SR)

PVFs	SR
$L = D - A$	$\Psi^\pi = \left(I - \gamma P^\pi\right)^{-1}$
$L\mathbf{v} = \lambda\mathbf{v}$	$\Psi^\pi\mathbf{v} = \mu\mathbf{v}$

When P^π is symmetric and π is the uniform random policy.

An Auxiliary Task Perspective on PVFs

$$\psi^\pi(s, a, s') = \mathbb{I}\{s = s'\} + \gamma \mathbb{E}_\pi\left[\psi^\pi(S_{t+1}, A_{t+1}, s'), \,\middle|\, S_t = s, A_t = a\right]$$



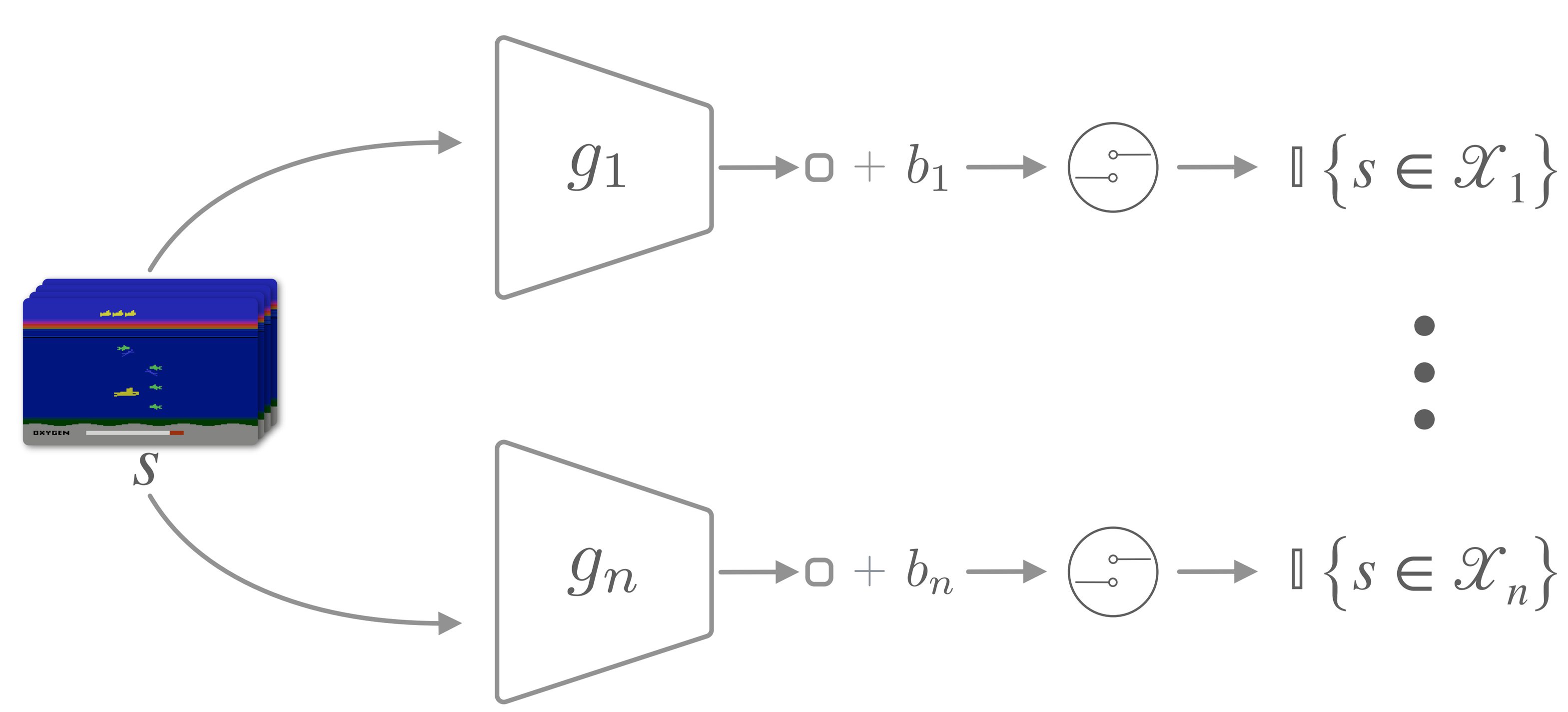
Training with auxiliary tasks produces representations Φ that span the principle components of Ψ^π (PVFs).

A Practical Implementation with the Successor Measure

- Generalize equality indicator to set-based indicator as motivated by the Successor Measure.

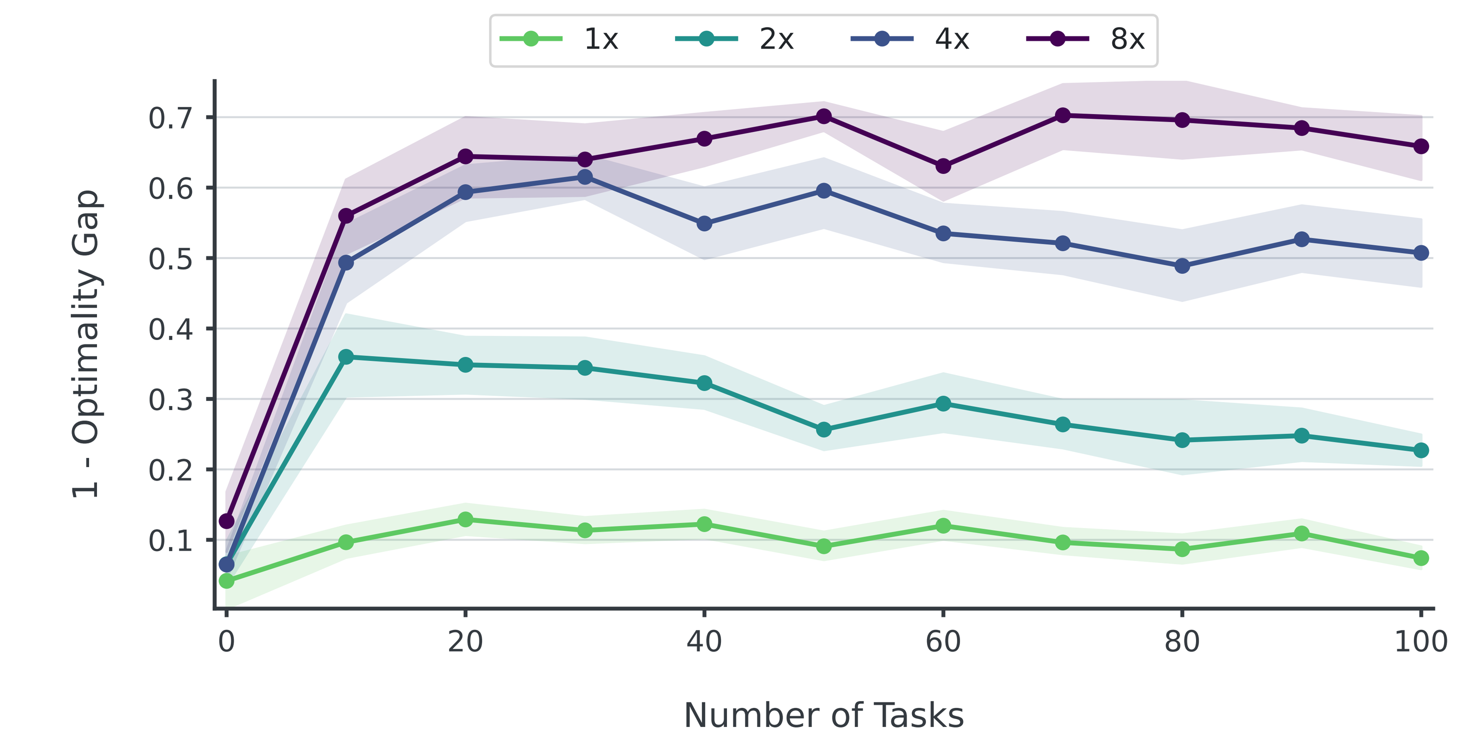
$$\psi^\pi(s, a, \mathcal{X}) = \mathbb{I}\{s \in \mathcal{X}\} + \gamma \mathbb{E}_\pi\left[\psi^\pi(S_{t+1}, A_{t+1}, \mathcal{X}), \,\middle|\, S_t = s, A_t = a\right]$$

Random Network Indicator

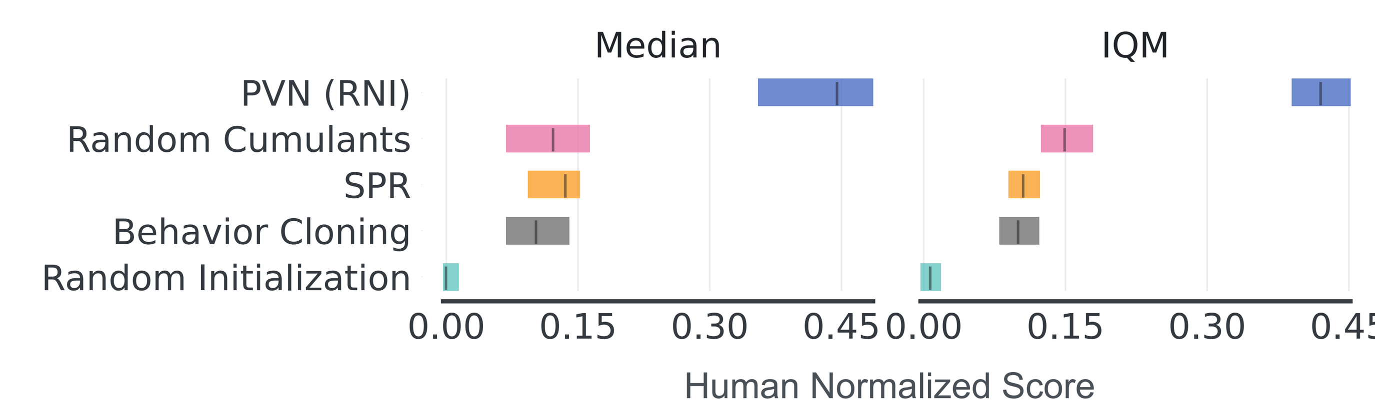


Evaluating PVNs via Online Linear Control

- When learning with many auxiliary tasks increasing capacity is crucial (specifically width).



- PVN learns a representation that's amenable to linear control with strong perf. in just 15M frames.



Qualitatively PVN Encodes Temporal Structure