Proper Value Equivalence

Niklas Forsstroem
forsstroemniklas@gmail.com

Paper: https://arxiv.org/abs/2106.10316

Introduction

What is the objective?

01 In model-based RL, we need to represent the environment

02 Modelling the entire world in not feasible due to complexity

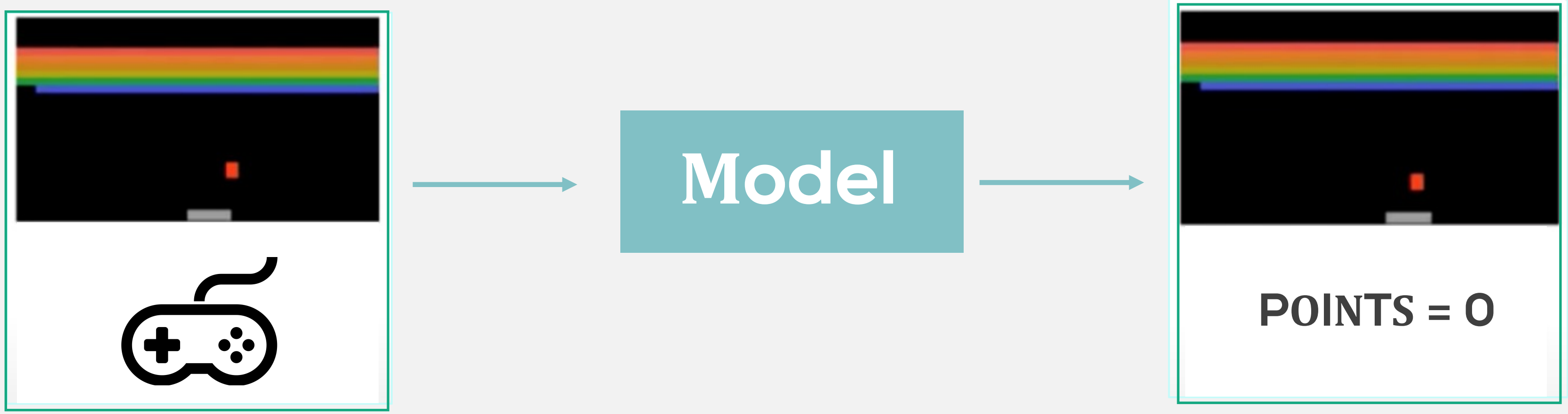
03 Want to reduce complexity whilst preserving necessary information for planning

Order-k Value Equivalent Class Definition

An order-k value equivalent class is defined as:

$$\mathcal{M}^k(\Pi, \mathcal{V}) \equiv \{\tilde{m} \in \mathcal{M} : \quad \tilde{\mathcal{T}}^k_\pi v = \mathcal{T}^k_\pi v \quad \forall \quad \pi \in \Pi, v \in \mathcal{V}\}$$

$\mathcal{M} \subseteq \mathbb{M}$ is a class of models and $\tilde{m}$ is a model



Bellman operator
Definition

The bellman operator $\mathcal{T}_\pi$ is defined as:

$$\mathcal{T}_\pi[v](s) \equiv \mathbb{E}_{A\sim\pi(\cdot|s),S'\sim p(\cdot|s,A)}[r(s,A) + \gamma v(S')]$$

$v \in \mathbb{V}$ is any function in the space $\mathbb{V} \equiv \{f \mid f: \mathcal{S} \mapsto \mathbb{R}\}$.

The operator implicitly depends on one's model (m) of the environment (e.g. via $S' \sim p(\cdot| s, A)$)

Can be shown that $lim_{n\to\infty} \mathcal{T}_\pi^n v = v_\pi$

starting from any $v \in \mathbb{V}$, repeated application of $\mathcal{T}_\pi$ will eventually converge to v_π.

Order-k value equivalent class

Definition

An order-k value equivalent class is defined as:

$$\mathcal{M}^k(\Pi, \mathcal{V}) \equiv \{\tilde{m} \in \mathcal{M} : \quad \tilde{\mathcal{T}}_\pi^k v = \mathcal{T}_\pi^k v \quad \forall \quad \pi \in \Pi, v \in \mathcal{V}\}$$

An extension of order-1 value equivalent class: $\mathcal{M}^1(\Pi, \mathcal{V}) \equiv \{\tilde{m} \in \mathcal{M} : \quad \tilde{\mathcal{T}}_\pi v = \mathcal{T}_\pi v \quad \forall \quad \pi \in \Pi, v \in \mathcal{V}\}$

$\tilde{\mathcal{T}}_\pi$ denotes one application of the Bellman operator induced by model $\tilde{m}$ and policy π to function v

$\mathcal{T}_\pi$ is the environment's Bellman operator for π.

Order-k value equivalent class

Illustrative example

Original model

Non-value equivalent model

Value equivalent model

Value equivalence
Restricting to k=1

Grimm et al. [1] studied order-one VE classes of the form

$$\mathcal{M}^1(\Pi, \mathcal{V}).$$

They have shown that $\mathcal{M}^1(\mathbb{\Pi}, \mathbb{V})$ either contains only the full environment or is empty.

As Π and $\mathcal{V}$ expand expand, the set of VE models shrinks, eventually collapsing to a single point corresponding to a perfect model.

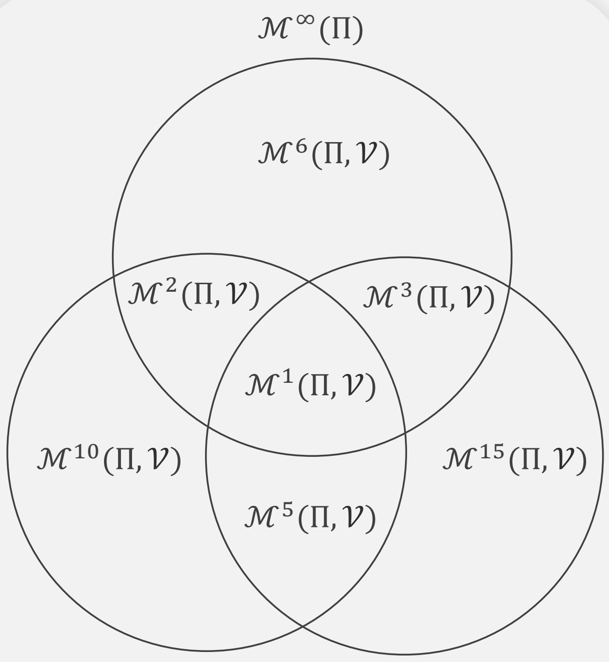
Would have to find trade-off between granular policies and functions vs simplistic permissible models

The above is not always the case for $k > 1$.

Proper value equivalence
Limit as $k \to \infty$

$\mathcal{M}^{\infty}$, the class of Proper Value Equivalent (PVE) models, is the set of models fulfilling the below:

$$\mathcal{M}^{\infty}(\Pi) = \lim_{k\to\infty} \mathcal{M}^{k}(\Pi, \mathbb{V}) = \{\tilde{m} \in \mathcal{M} \ : \ \tilde{v}_{\pi} = v_{\pi} \ \ \forall \pi \in \Pi\}$$



Since repeatedly applying $\tilde{\mathcal{T}}_{\pi}$ to a function $\tilde{v}_{\pi}$ converges to the same fixed point regardless of the function $\tilde{v}_{\pi}$, in an order-∞ VE class the set Π uniquely determines the set $\mathcal{V}$

$\mathcal{M}^{\infty}$ is the "biggest" VE class. One can also define this special VE class in terms of any other:

$$\mathcal{M}^{\infty}(\Pi) = \bigcap_{\pi \in \Pi} \mathcal{M}^{k}(\{\pi\}, \{v_{\pi}\})$$

Proper value equivalence

Why $\mathcal{M}^\infty$ is useful

Unlike with VE, the class of PVE models does not collapse in the policy / function limit.

Even if all value functions are used, we generally end up with multiple PVE models.

All of these models are sufficient for planning, meaning that they will yield an optimal policy

Some of these are easier to learn or represent than others.