Learning Markov State Abstractions for Deep Reinforcement Learning

Cameron Allen, Neev Parikh, Omer Gottesman, George Konidaris — Brown University

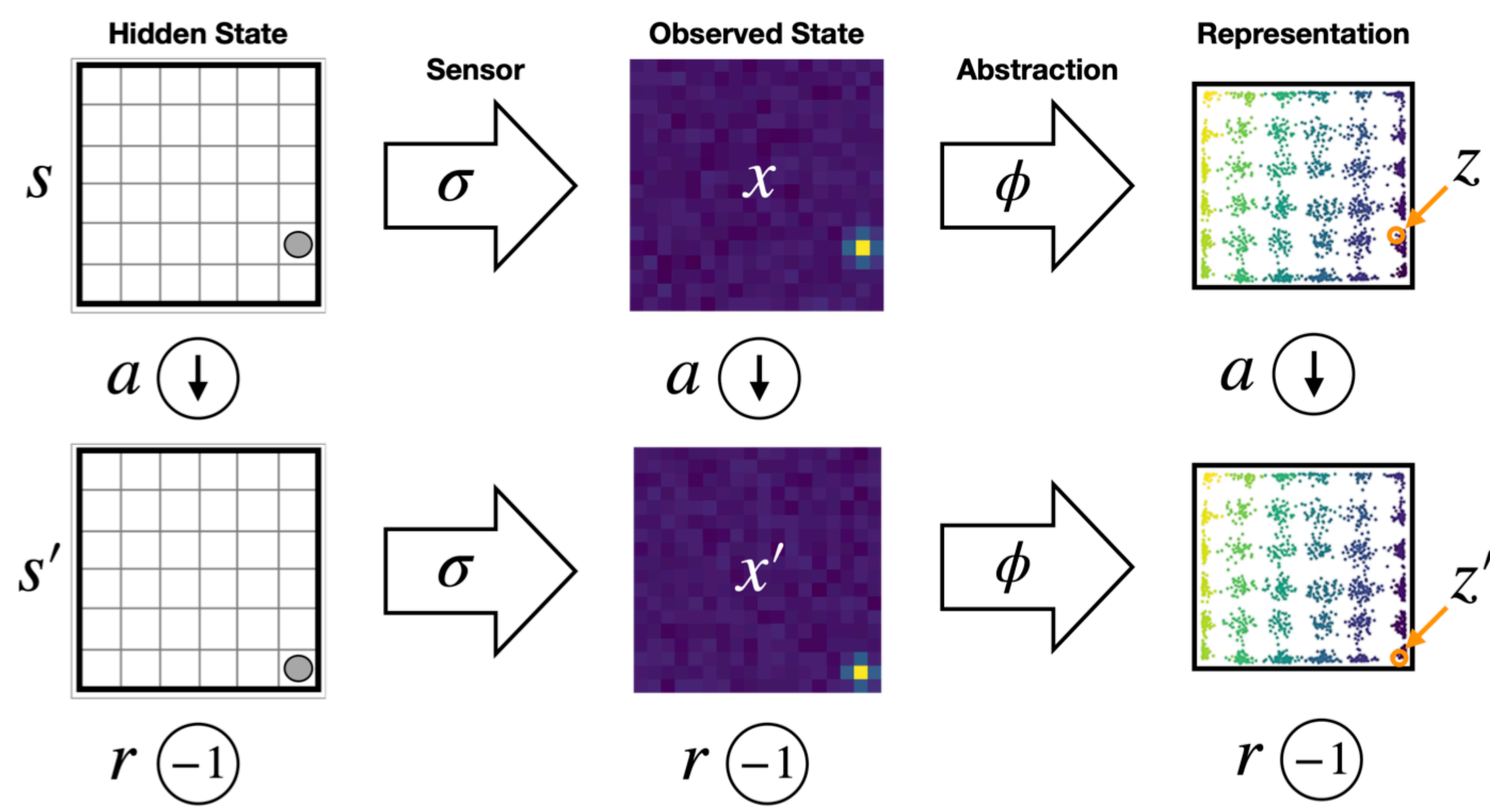
csal@brown.edu

@camall3n

The Markov Property

A decision process is ***Markov*** if each state x is a sufficient statistic, given any action a, for predicting the distribution over next states x' and the expected reward r —no additional history is required.

$$T(x' \mid a, x) = T(x' \mid a, x, \{a_{t-1}, x_{t-1}, \cdots\})$$
$$R(x', a, x) = R(x', a, x, \{a_{t-1}, x_{t-1}, \cdots\})$$

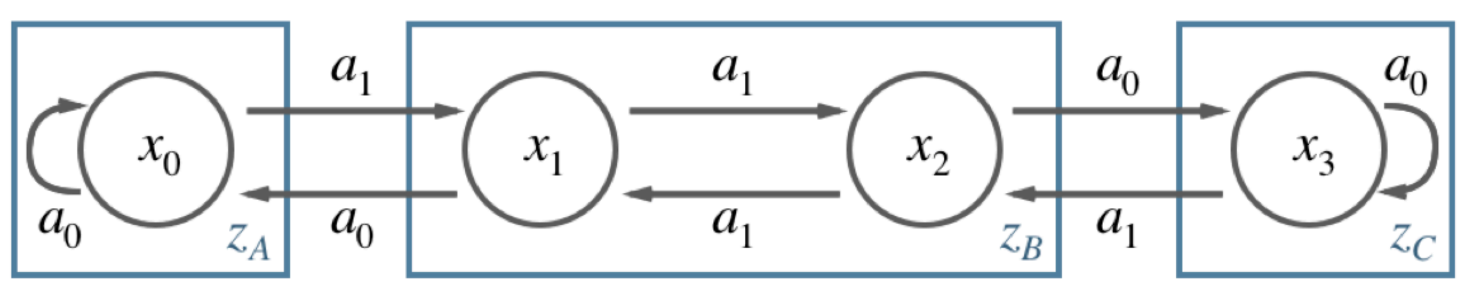


Training a model to both **discriminate** and **explain** state transitions encourages **Markov** abstract states that **improve RL** performance.

State Abstraction

An **abstraction** $\phi : X \to Z$ maps ground states x to abstract states $z = \phi(x)$, with the hope that learning is more tractable in Z.

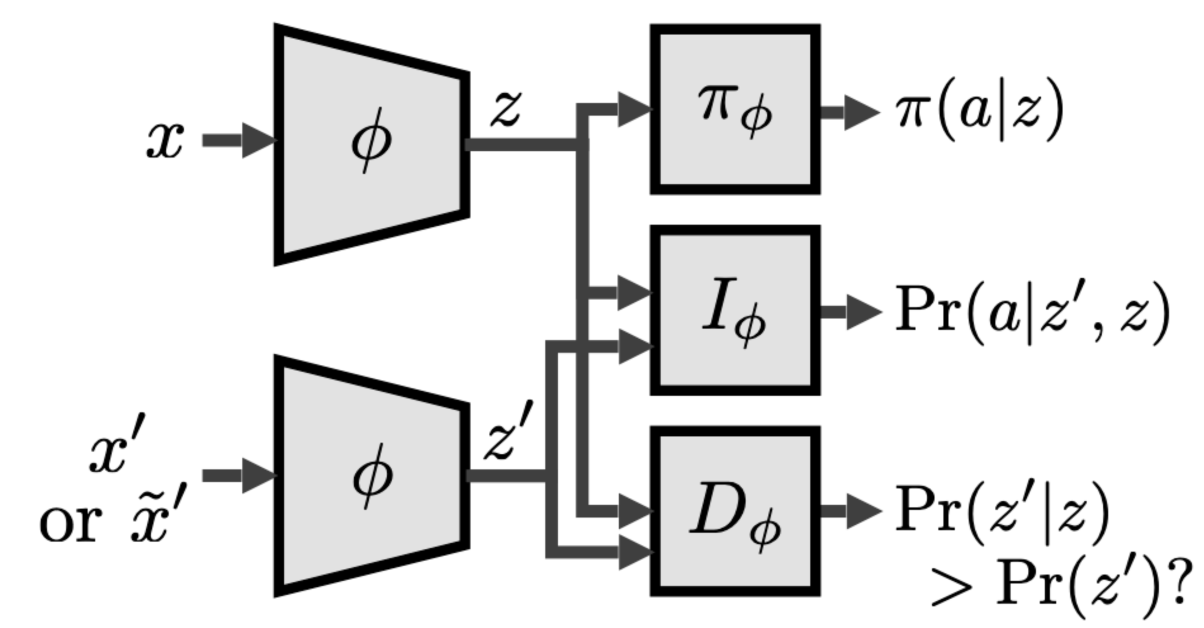
Any abstraction ϕ, when applied to an MDP M, induces a new abstract MDP $M_\phi = (Z, A, T^\pi_{\phi,t}, R^\pi_{\phi,t}, \gamma)$, whose dynamics may depend on the current time step t, or the agent's behavior policy π, and crucially, which **is not guaranteed to be Markov**.



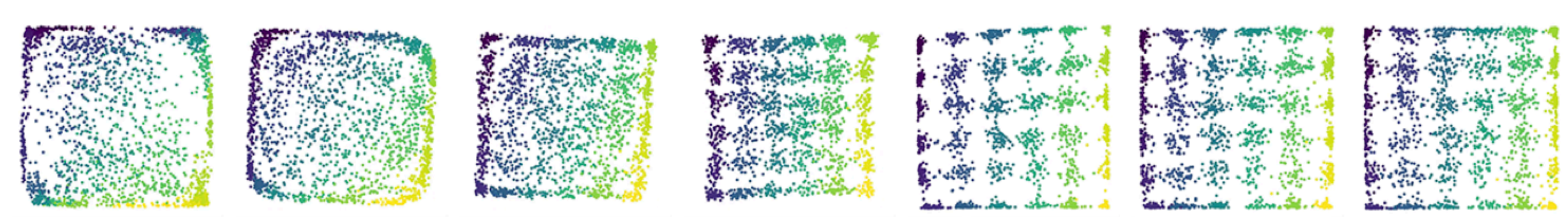
Example: An MDP and a non-Markov abstraction. The abstract transition probabilities depend on history beyond just the most recent abstract state.

System Architecture

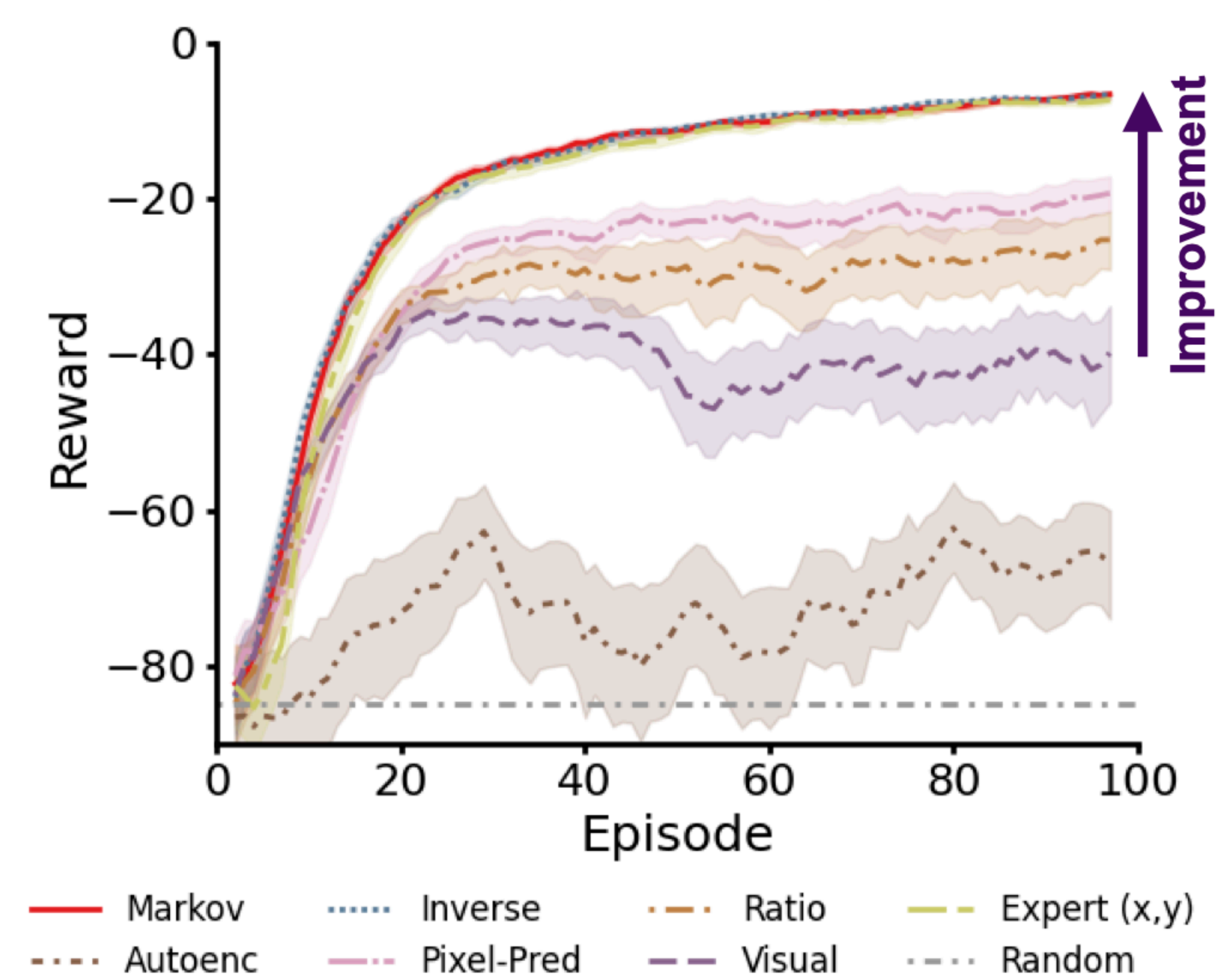
An encoder ϕ maps ground states x to abstract states z, which are then used as inputs for an abstract state-transition discriminator and an inverse dynamics model. The agent's policy π depends only on the abstract state.



Representation Learning Progress of a 2-D Markov state abstraction for a 6×6 visual gridworld. Color denotes ground-truth (x, y) position (not shown to agent).



Gridworld Results. Markov state abstractions **fully close** the representation gap, matching expert (x,y).



Theorem 1: Markov State Abstractions

Given an MDP and an abstraction ϕ, if the following **conditions** hold, for any abstract policy:

1. The ground and abstract state transitions have equal **inverse model** probabilities
$$\Pr(a|z', z) = \Pr(a|x', x)$$
2. The ground and abstract state transitions have equal next-state **density ratios**
$$\frac{\Pr(z'|z)}{\Pr(z')} = \frac{\Pr(x'|z)}{\Pr(x')}$$

Then ϕ is a **Markov abstraction**.

DeepMind Control Suite Results. Markov state abstractions lead to **state-of-the-art** learning performance over baseline **RAD**.