POMDP Homomorphisms

Alicia Peregrin Wolfe

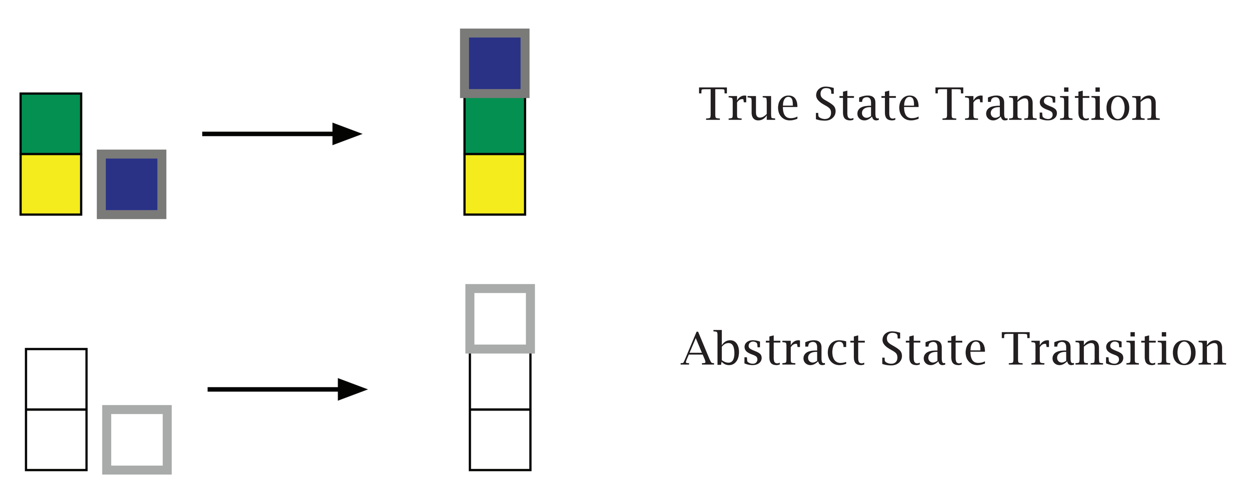
Autonomous Learning Laboratory
University of Massachusetts, Amherst
pippin@cs.umass.edu

Abstract

The problem of finding hidden state in a POMDP and the problem of finding state abstractions for MDPs are closely related. This work analyzes the connection between existing Predictive State Representation methods and homomorphic reductions of Markov Processes. We formally define a POMDP homomorphism, then extend PSR reduction methods to find POMDP homomorphisms when the original POMDP is known. The resulting methods find more compact abstract models than PSR reduction methods in situations where different observations have the same meaning for some task or set of tasks.

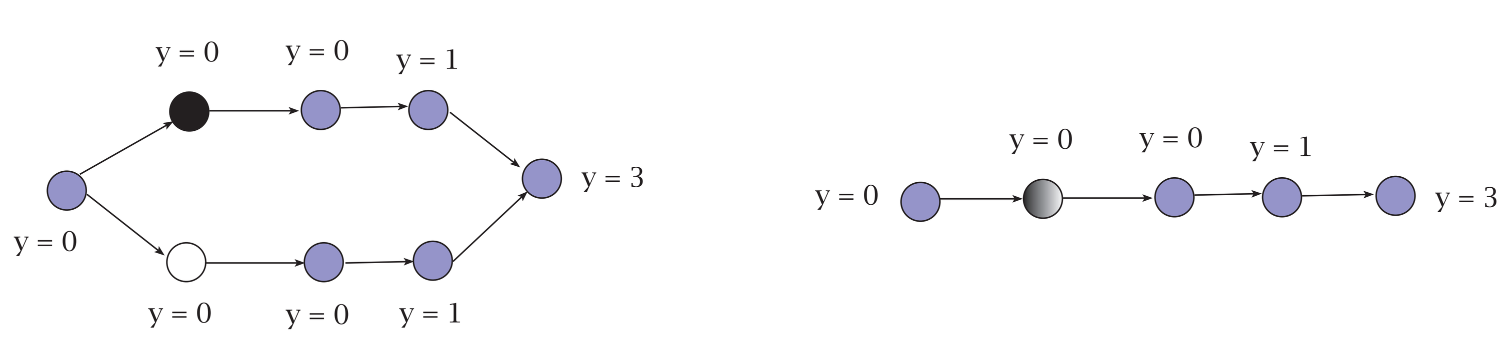
Model Minimization

Find a smaller model which maintains only the relevant properties of the original model, with respect to some output variable *y*.



POMDP Homomorphisms

Reduction over states, actions and observations

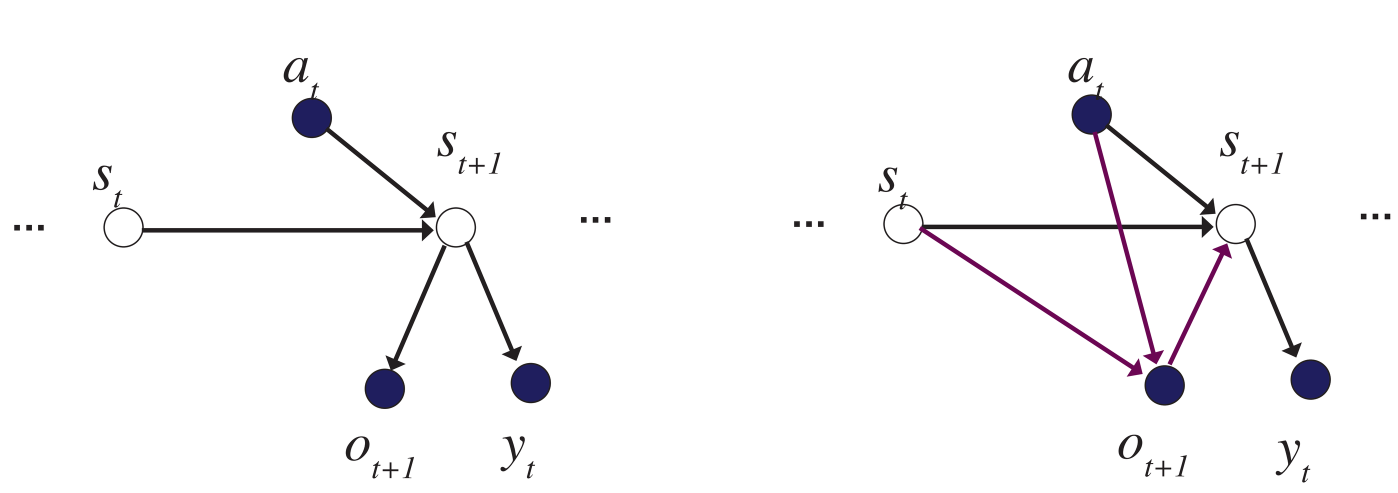


State, action and observation mappings:

$f: S \rightarrow S'$ $\quad g: A \rightarrow A'$ $\quad k_a: O \rightarrow O'$

Seek to predict some specific output variable *y*, where *y* is a function of the observation.

Constraints (Bayes Net View)



Output:

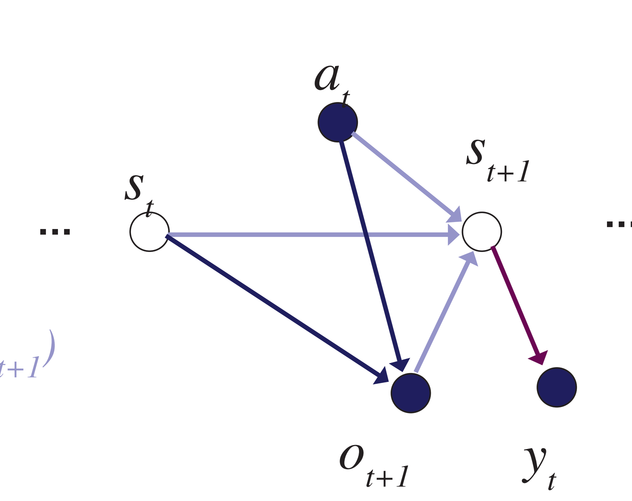
- $P(y_t \mid f(s_t)) = P(y_t \mid s_t)$

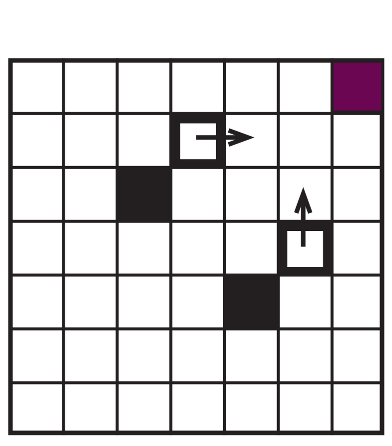
Transitions:

- $P(f(s_{t+1}) \mid f(s_t), g(a_t), k_a(o_{t+1})) = P(f(s_{t+1}) \mid s_t, a_t, o_{t+1})$

Observations:

- $P(k_a(o_{t+1}) \mid f(s_t), g(a_t)) = P(k_a(o_{t+1}) \mid s_t, a_t)$





State Specific Action/Observation Mappings

If agent could believe that it might be in s_1 or s_2, cannot have different action mappings in those states.

History specific action/obsevation mappings may be easier.

Linear PSR Algorithm

- Tests: $t = a_1o_1a_2o_2a_3o_3$, $P(t \mid s) = P(o_1o_2o_3 \mid sa_1a_2a_3)$
- State represented by set of linearly independent tests: $q_i \in Q$
- State Mapping: $f(s_1) = f(s_2) \Leftrightarrow \forall\, q_i\; P(q_i \mid s_1) = P(q_i \mid s_2)$

	q_1	...	q_2	...
s_1	0.3		0.2	
s_2	0.4		0.5	
s_3	0.3		0.2	
...				

- If m_t is the prediction for test *t*, update vector for state consists of: m_{aoq_i}/m_{ao} for all tests q_i
- Action Mapping: $g(a), k(o) = g(a'), k(o') \Leftrightarrow \forall\, q_i\; m_{aoq_i}/m_{ao} = m_{a'o'q_i}/m_{a'o'}$

Output Function (y) Homomorphisms

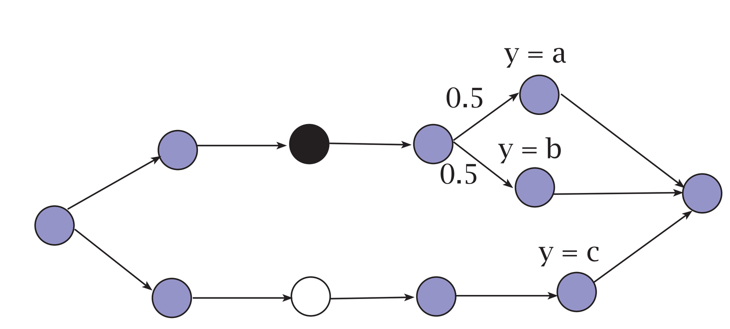
1. Initial set of tests: a_1y_1 (one time step, *y* observed)
2. Split *a,o* pairs which help predict *Q*
3. Extend tests by one time step using *g(a), k(o)*
4. Repeat (2, 3) until no change

Value Function Homomorphims

Start with the immediate reward as the only basis vector, as in (Poupart, 2002).

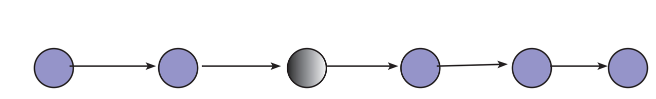
Results

Original POMDP:

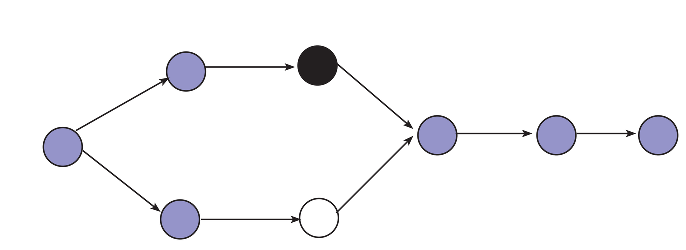


Task one: a = b = c

Homomorphic reduction

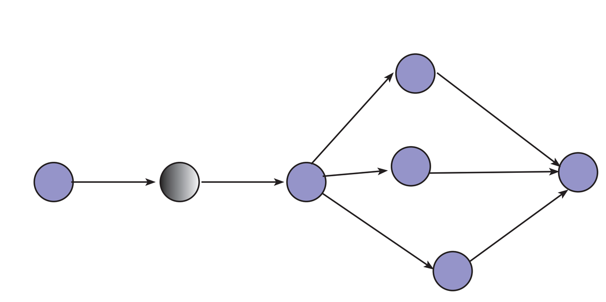


PSR Reduction

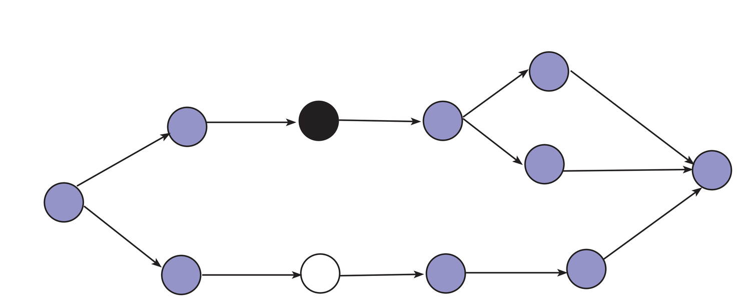


Task two: a = 2, b = 1, c = 1.5 ((a+b)/2 = c)

Value Function Reduction



Output Function Reduction



Acknowledgements

This research was facilitated in part by a National Physical Science Consortium Fellowship and by stipend support from Sandia National Laboratories, CA. This research was also funded in part by NSF grant CCF 0432143.

Citations

Thomas Dean and Robert Givan. Model minimization in markov decision processes. AAAI, 1997.

Masoumeh T. Izadi and Doina Precup. Model minimization by linear psr. IJCAI, 2005.

Michael L. Littman, Richard S. Sutton, and Satinder P. Singh. Predictive representations of state. NIPS, 2001.

Pascal Poupart and Craig Boutilier. Value-directed compression of pomdps. NIPS, 2002.

B Ravindran. An Algebraic Approach to Abstraction in Reinforcement Learning. PhD thesis, University of Massachusetts, 2004.

Alicia Peregrin Wolfe and Andrew G. Barto. Decision tree methods for finding reuseable mdp homomorphisms. AAAI, 2006.