

Generating Exponentially Smaller POMDPs with Conditionally Irrelevant Variable Abstraction

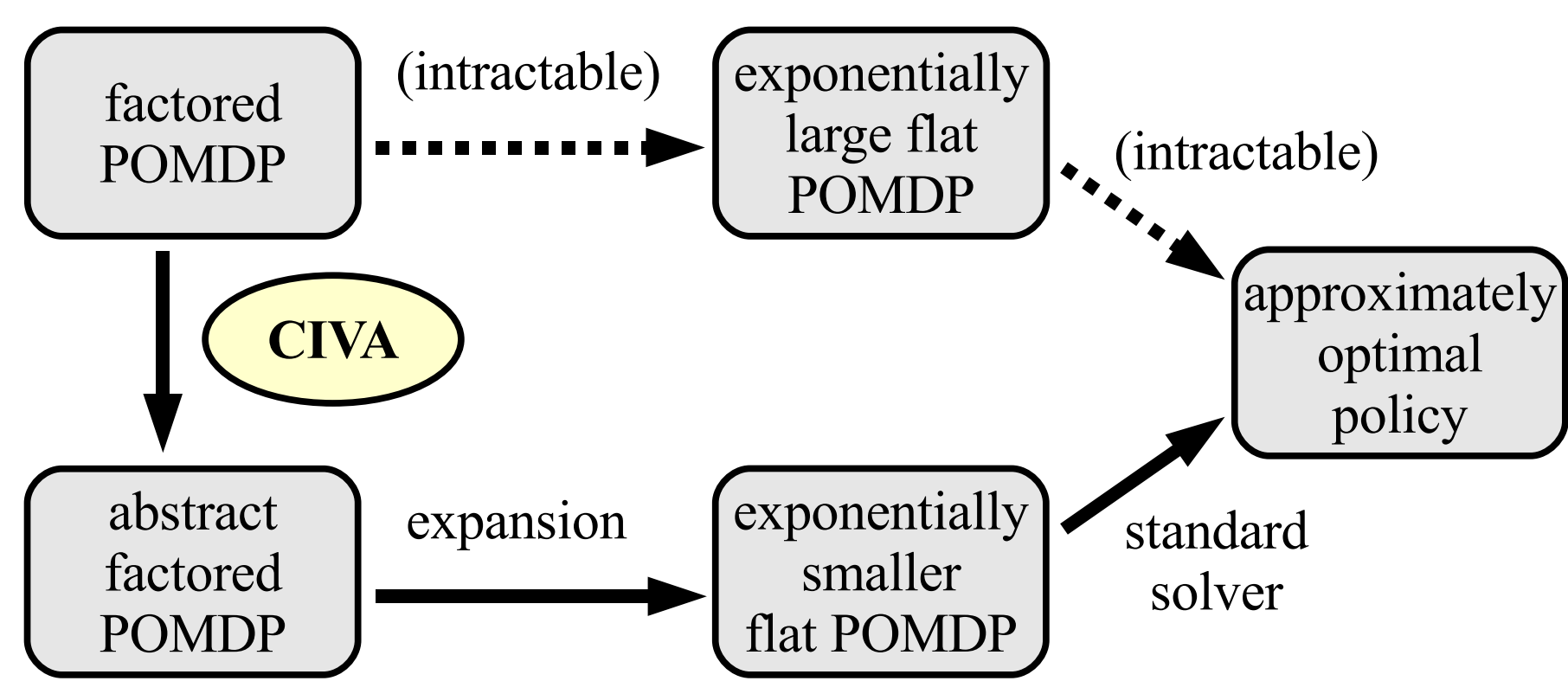
Trey Smith
Carnegie Mellon West / NASA Ames

David R. Thompson and David Wettergreen
Carnegie Mellon Robotics Institute

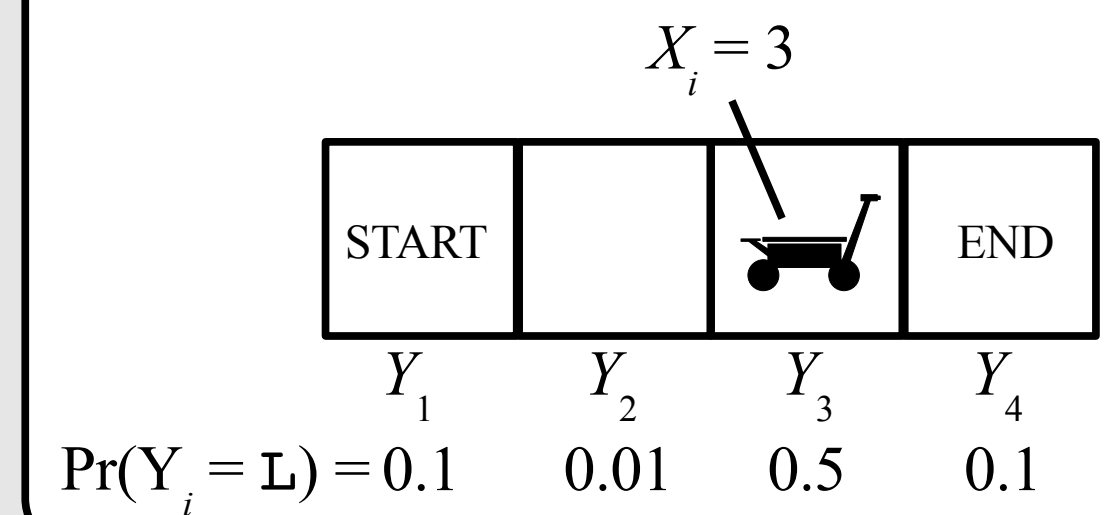
Abstract

The state of a POMDP can often be factored into a tuple of n state variables. The corresponding flat model, with size exponential in n , may be intractably large. We present a novel method called conditionally irrelevant variable abstraction (CIVA) for losslessly compressing the factored model, which is then expanded into an exponentially smaller flat model in a representation compatible with many existing POMDP solvers. We applied CIVA to previously intractable problems from a robotic exploration domain. We were able to abstract, expand, and approximately solve POMDPs that had up to 10^{24} states in the uncompressed flat representation.

CIVA Process Diagram



MiniLifeSurvey Example Problem



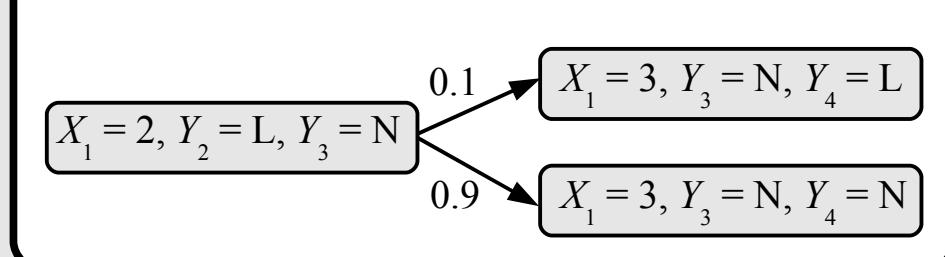
MiniLifeSurvey Actions

- move**: moves the robot one cell forward
- scan**: applies a sensor to the forward cell, returning a noisy reading as to whether it contains life
- sample**: returns a reward if the current cell contains life

MiniLifeSurvey Variables

- X_i : position, ranges 1..k
- $Y_1..Y_k$: binary, Y_i indicates life (L) or no life (N) in cell i

Abstract State Transition



- When $X_i=2$, only Y_2 and Y_3 are conditionally relevant
- When $X_i=3$, only Y_3 and Y_4 are conditionally relevant
- Naive flat model contains $k \times 2^k$ states
- After CIVA, only $4k$ states

POMDP Review and Notation

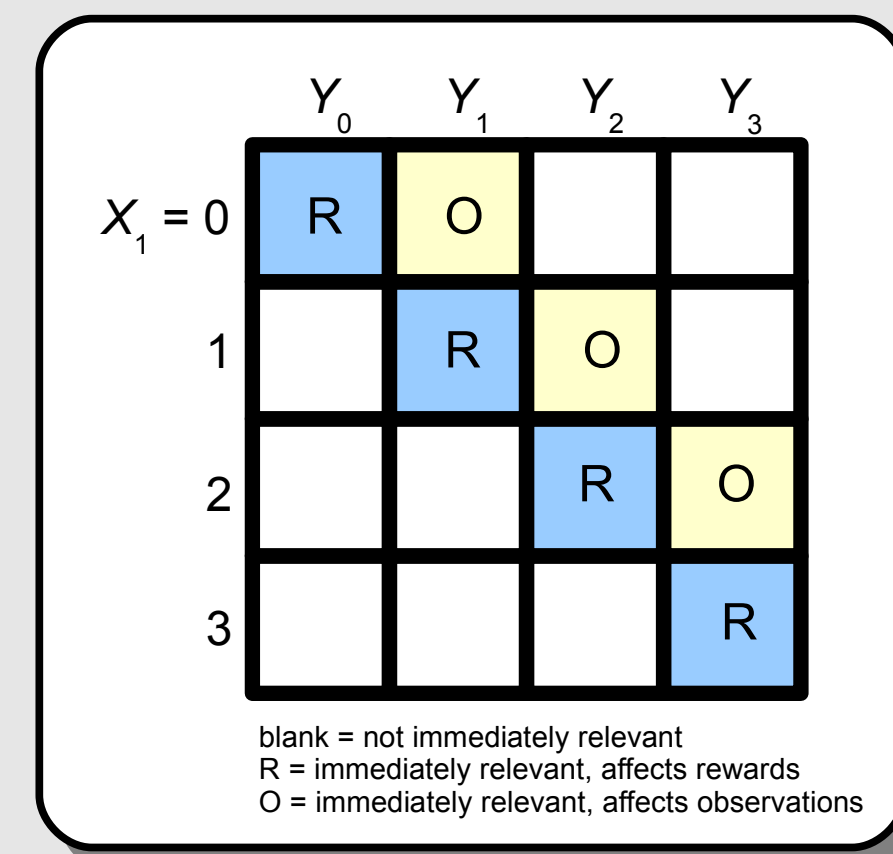
- $P = \langle S, A, T, R, \gamma, O \rangle$
- S set of states, A set of actions, $T(s,a,s')$ probability of transitioning from s to s' given action a is applied, $R(s,a)$ reward for applying action a in state s , γ discount factor, $O(a,s',o)$ chance of seeing observation o when applying action a and transitioning into state s' .
- $a_i = \text{pi}(a_{o'} o_{o'} a_{r'} o_{r'} \dots a_{t'} o_{t'})$
- Optimize expected sum of discounted rewards
- Alternately, $a_i = \text{pi}(b_i)$
- $P = \langle S, A, T, R, \dots \rangle$, $P' = \langle S', A', T', R', \dots \rangle$
- P, P' **policy-compatible** if $A = A', O = O'$ -- can use policy from P on P' , same functional form
- P, P' **policy-equivalent** if all policies have the same expected reward across both problems
- CIVA takes a POMDP P and produces a policy-equivalent POMDP P'

Upstream and Downstream

- Upstream variables**: $X_1..X_k$ always known, transition deterministically
- Downstream variables**: $Y_1..Y_k$ all other variables, need not transition deterministically, do not affect upstream variables
- Variables manually labeled as upstream or downstream
- Upstream variables always relevant, downstream variables may be conditionally irrelevant depending on the values of the upstream variables

Immediately Relevant

- Y_i **immediately relevant** at x if:
 - Y_i affects immediate rewards, and
 - Y_i affects immediate observations, and
 - Y_i affects transition of other vars

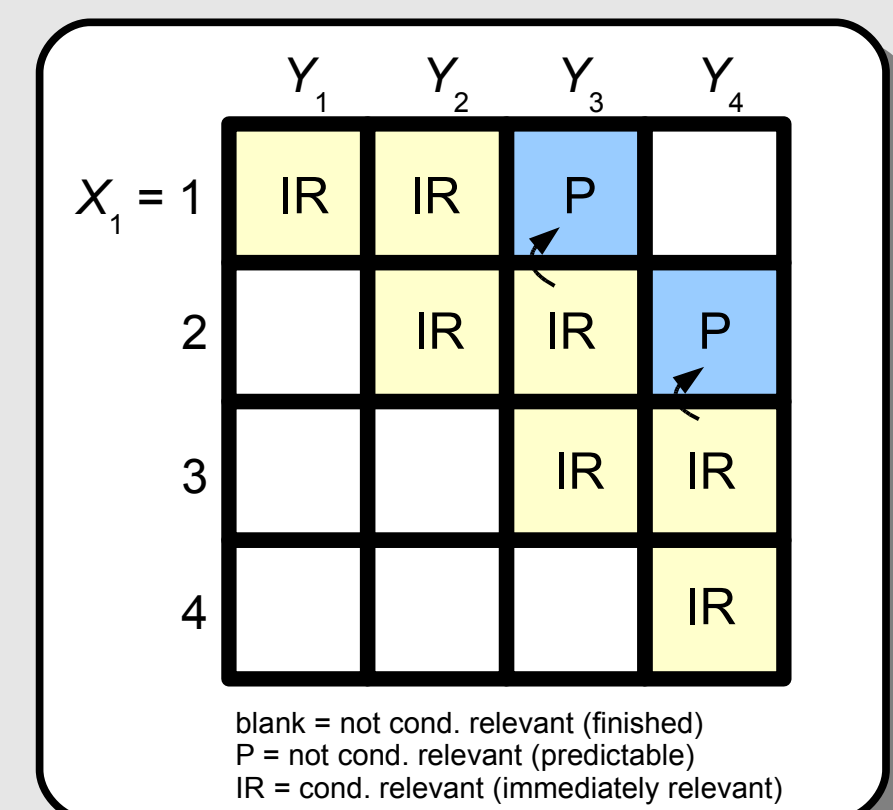
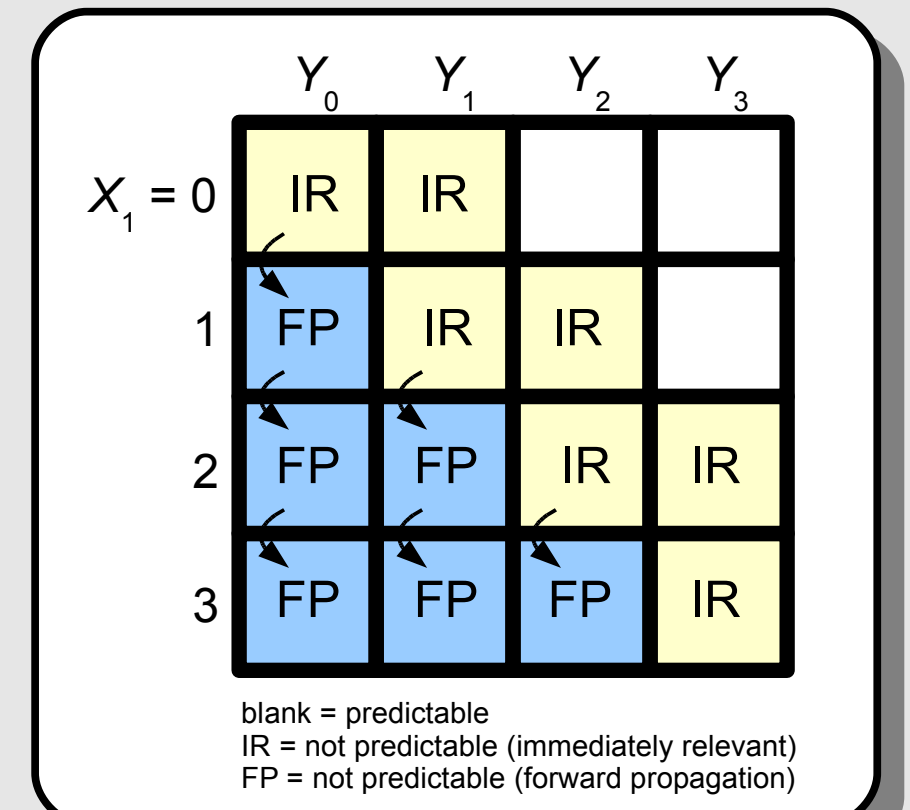


Conditionally Relevant

- Y_i **conditionally relevant** at x if:
 - Y_i immediately relevant at x , or
 - Y_i immediately relevant at successor $x' = U(x,a)$ and not predictable after (x,a)

Predictable (Untouched)

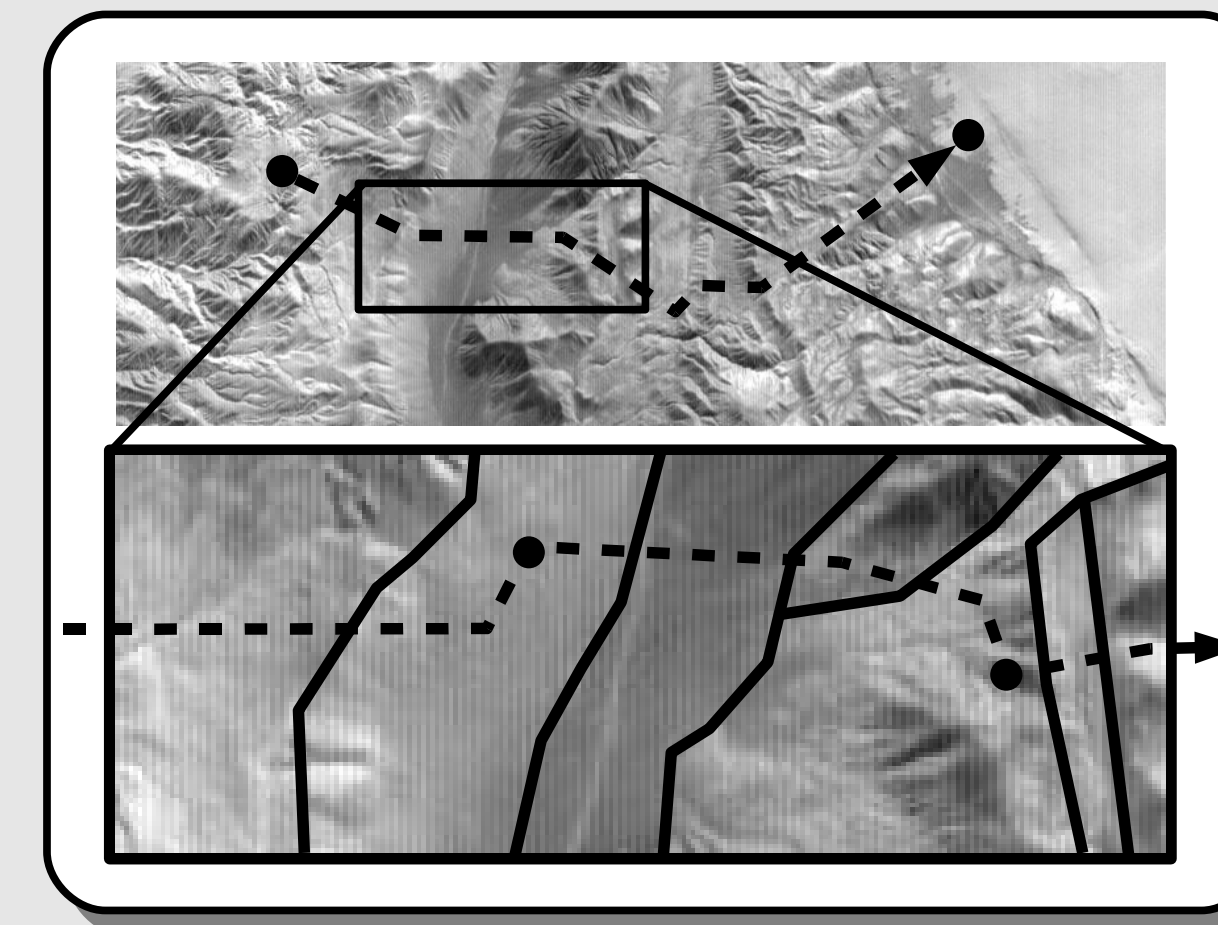
- Y_i **predictable (untouched)** after (x,a) if:
 - Y_i independent of other vars in initial belief, and
 - Y_i value does not change over time, and
 - Y_i not immediately relevant at x , and
 - Y_i is untouched after predecessors x' of x



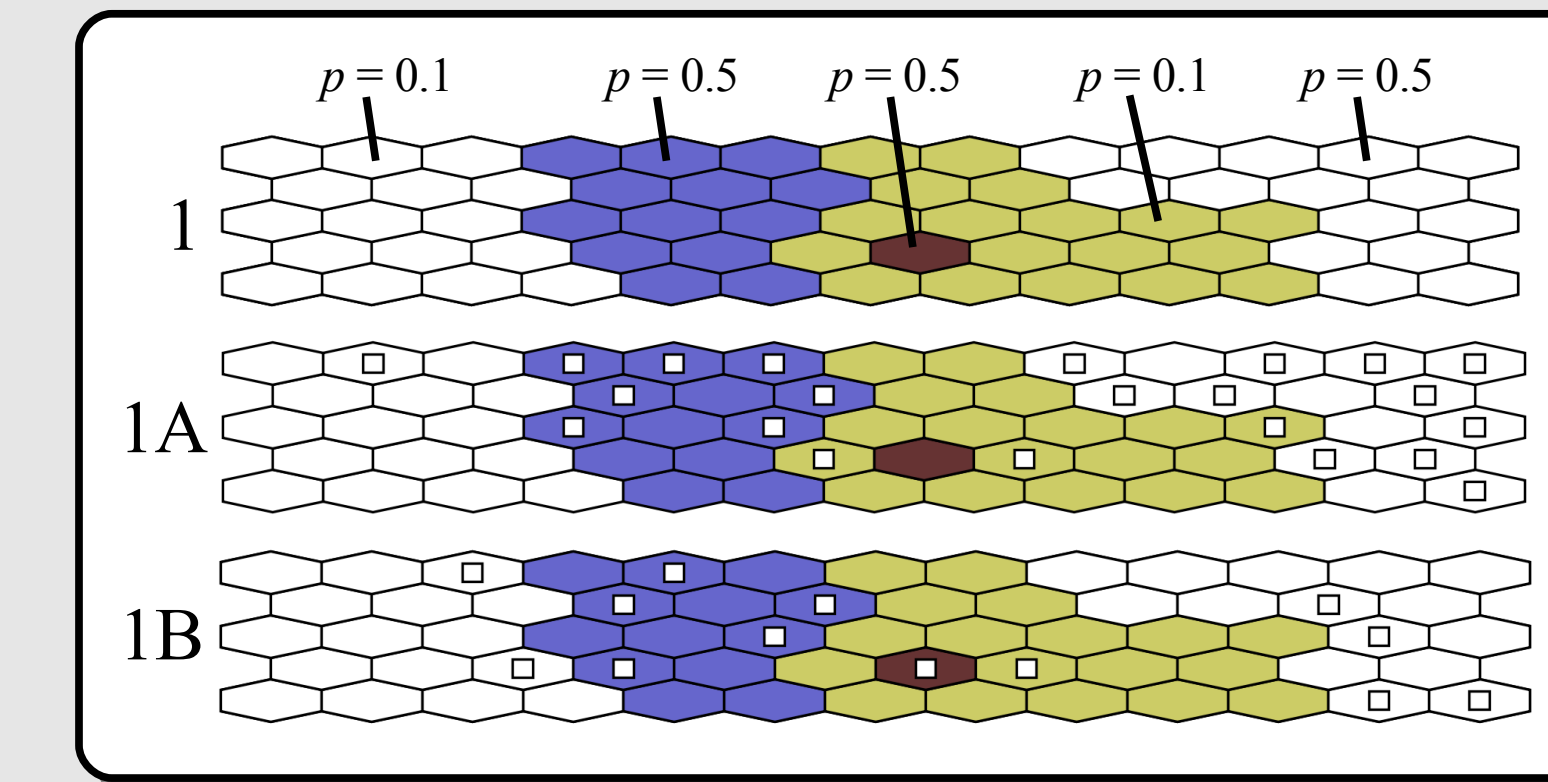
Related Work

- Boutelier and Dearden '94**: Globally irrelevant variables for MDPs
- Baum and Nicholson '98**: Non-uniform abstraction depending on planning horizon
- Solve factored model directly:
 - McAllester and Singh '99**: Boyen-Koller belief simplification
 - St. Aubin et al. '00**: ADD representation of MDPs
 - Hansen and Feng '00**: Extended ADDs to POMDPs with factored observations

Full LifeSurvey Problem



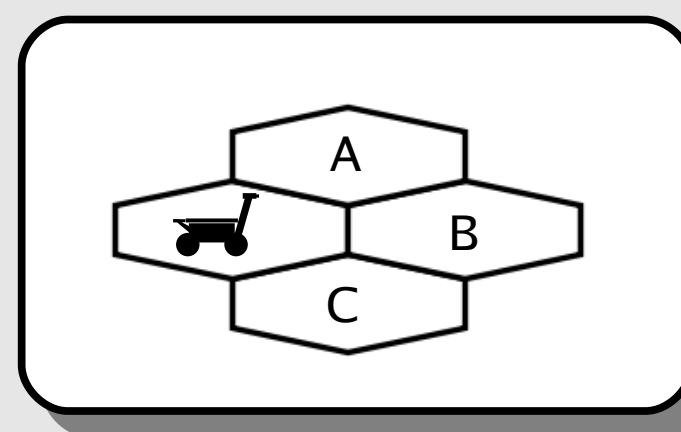
Traverse planning at multiple scales



Prior map 1 and randomly drawn target layouts used for testing

Available Actions

- move**: move to A, B, or C (cost 1)
- sampling move**: sample A, B, or C (cost 5)
- scan**: scan all three cells (cost 5)



The Zoë rover at the test site in Pittsburgh

Per-Region Reward

- +50 if sampled cell w/life, or
- +20 if passed through cell w/life, or
- +5 if region was entered

Scan Results

- A single scan action returns three independent noisy readings from the forward cells
- Each reading has three possible values "negative", "maybe", "positive"
- Observation model learned from data collected by robot
- Distribution with life: 72% / 12% / 16%
- Distribution without life: 9% / 5% / 86%

CIVA Results

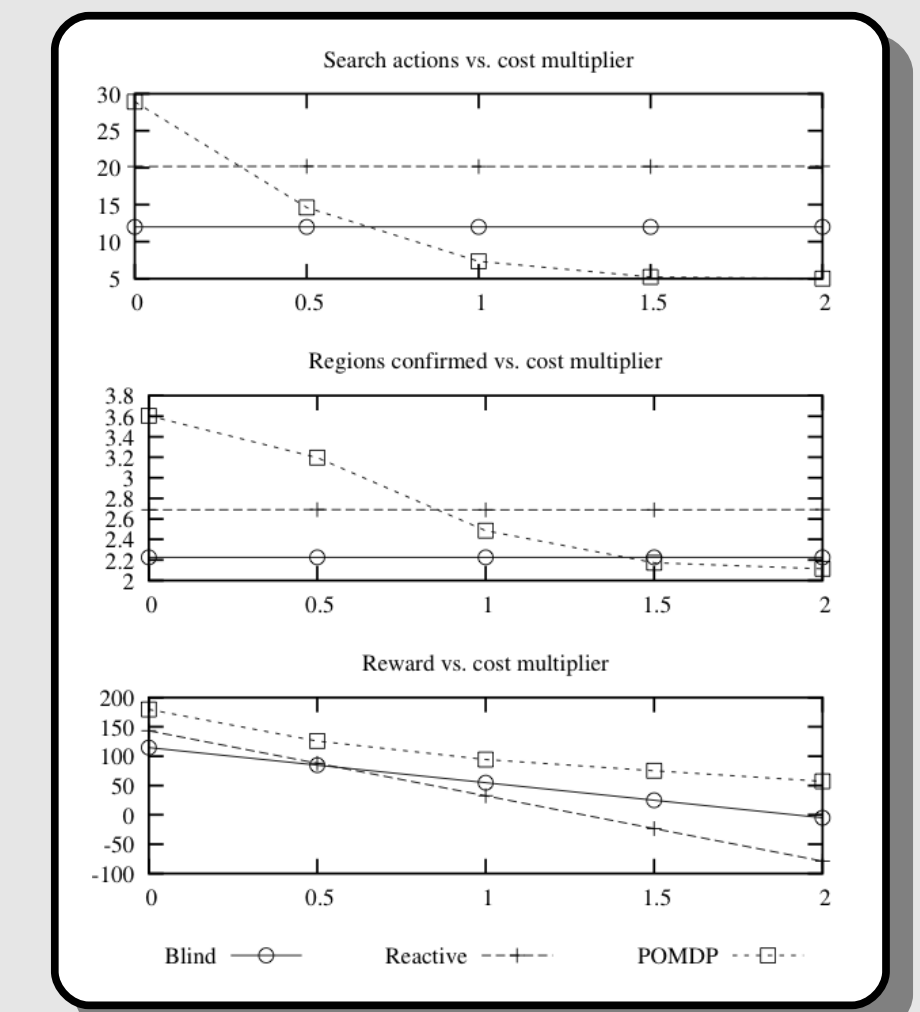
- Prior map 1: 3.57×10^{24} states \rightarrow 7,001
- Prior map 2: 1.43×10^{25} states \rightarrow 7,841
- CIVA compression took < 2 seconds
- Ran ZMDP solver for about 20 minutes
- Tested resulting plan over 20 runs in the outdoor environment

Onboard Testing

Planner	Search actions	Regions confirmed	Reward
Blind	12.0 ± 0.0	2.5 ± 0.0	68 ± 0
Reactive	20.0 ± 0.7	3.4 ± 0.6	61 ± 19
POMDP	7.5 ± 1.0	3.0 ± 0.5	113 ± 16

Adaptation to Changes

- Adjusted ratio of costs to rewards: $c=1$ is the original problem, $c=2$ high costs, $c=0$ no costs
- POMDP planner adjusts strategy in interesting ways
- General trend to reduce search actions (receiving less reward) as c increases



Conclusions

- Technique for abstracting away variables that are only *sometimes* irrelevant
- Exponentially smaller flat model
- Relies on fairly special structure -- "forward progress" assumption
- More generally brings up the idea of folding initial belief / reachability information into the transition function
 - Simpler transition function
 - More state aggregation

Acknowledgments

Thanks to the Life in the Atacama Project team for building and maintaining Zoe, and especially to Dominic Jonak for assisting in field experiments. This research was supported by NASA under grants NNG0-4GB66G and NAG5-12890.