Search and Explore: Symbiotic Policy Synthesis in POMDPs

Roman Andriushchenko[1] Alexander Bork[2] Milan Češka[1]
Sebastian Junges[3] Joost-Pieter Katoen[2] **Filip Macák[1]**

[1]**Brno University of Technology, Brno, Czech Republic**

[2]RWTH Aachen University, Aachen, Germany

[3]Radboud University, Nijmegen, The Netherlands

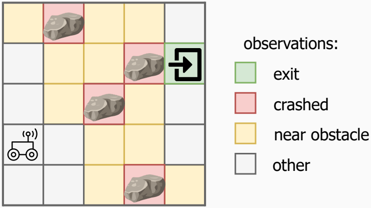
Published in **CAV'23**

Based on my master's thesis:
"Improving Synthesis of Finite State Controllers for POMDPs Using Belief Space Approximation"

Introduction

Partially-observable Markov decision processes (**POMDPs**)

- prominent model for sequential decision-making under uncertainty and limited observability
- observations – states with the same observation are indistinguishable
- observation-based polices are required



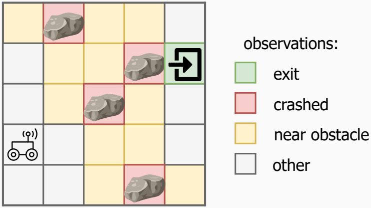
Specification:

- minimise the number of steps to reach the exit
- keep the probability of crashing below 1%

Introduction

Partially-observable Markov decision processes (**POMDPs**)

- prominent model for sequential decision-making under uncertainty and limited observability
- observations – states with the same observation are indistinguishable
- observation-based polices are required



Specification:

- minimise the number of steps to reach the exit
- keep the probability of crashing below 1%

Many practical applications:

- planning of autonomous agents and robotics
- games with imperfect information (e.g texas holdem)
- medical treatment strategies (e.g heart disease)

Problem Formulation

Find the optimal policy for the given **indefinite-horizon specifications**

- no discounting – much harder than finite-horizon problems
- important for long-term planning and sparse-rewards problems
- in general **undecidable** – policy may require infinite memory

Problem Formulation

Find the optimal policy for the given **indefinite-horizon specifications**

- no discounting – much harder than finite-horizon problems
- important for long-term planning and sparse-rewards problems
- in general **undecidable** – policy may require infinite memory

We aim at compact, verifiable and easy-to-execute strategies

- **finite-state controller** (**FSC**) based on Mealy machines

Problem Formulation

Find the optimal policy for the given **indefinite-horizon specifications**

- no discounting – much harder than finite-horizon problems
- important for long-term planning and sparse-rewards problems
- in general **undecidable** – policy may require infinite memory

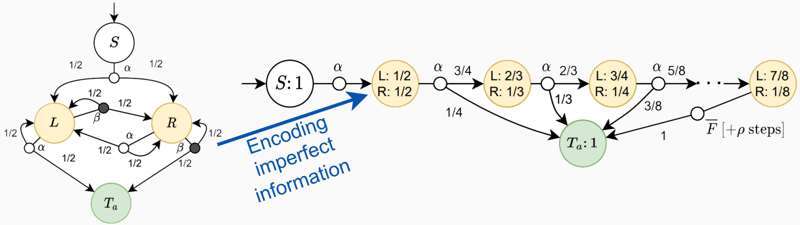
We aim at compact, verifiable and easy-to-execute strategies

- **finite-state controller** (**FSC**) based on Mealy machines

Anytime algorithm: in the given time, find the best FSC

SoTA I: Belief-based Methods

Belief - probability distribution over the states of a POMDP

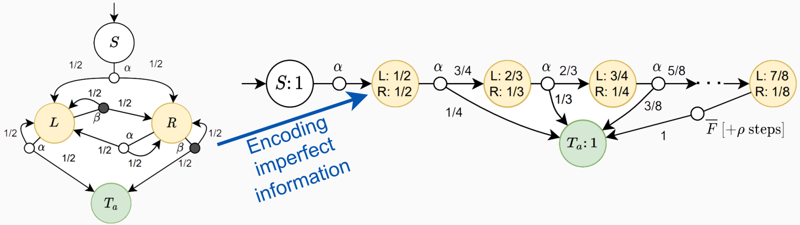


Construct and analyse the reachable belief space (i.e. belief MDP)

- belief MDPs are typically huge or even infinite
- various approximations of the unexplored belief space: **cut-offs** (e.g. in Storm) and **point-based approximations** (e.g. in SARSOP)

SoTA I: Belief-based Methods

Belief - probability distribution over the states of a POMDP



Construct and analyse the reachable belief space (i.e. belief MDP)

- belief MDPs are typically huge or even infinite
- various approximations of the unexplored belief space: **cut-offs** (e.g. in Storm) and **point-based approximations** (e.g. in SARSOP)

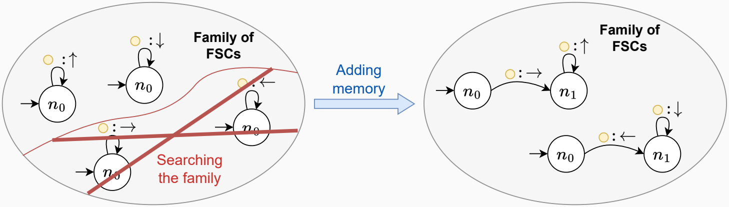
Limitations:

- cut-offs may provide very imprecise bounds or do not reduce the belief-space sufficiently
- point-based methods typically perform poorly for long-term planning

SoTA II: Inductive Synthesis of FSCs

Symbolic representation and exploration of families of candidate FSCs

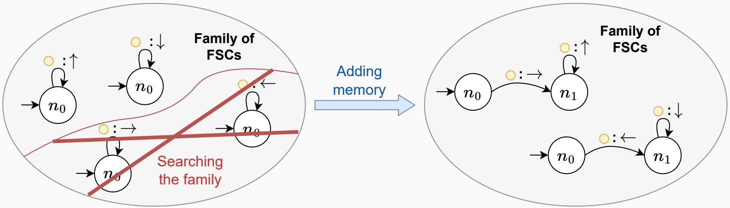
- fully-observable abstraction and counter-examples steer the exploration
- iterative expansion of the family by adding memory nodes
- implemented in the tool **PAYNT**



SoTA II: Inductive Synthesis of FSCs

Symbolic representation and exploration of families of candidate FSCs

- fully-observable abstraction and counter-examples steer the exploration
- iterative expansion of the family by adding memory nodes
- implemented in the tool **PAYNT**



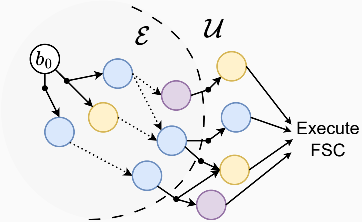
Limitations:

- the family size grows exponentially with the memory
- if a lot of memory is needed or the POMDP is too large, exploration becomes computationally intractable

Two Main Ideas of Our Approach

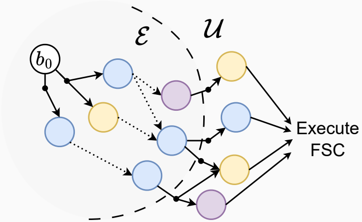
Two Main Ideas of Our Approach

Using FSCs as cut-offs to obtain a better approximation of the unexplored parts of the belief space



Two Main Ideas of Our Approach

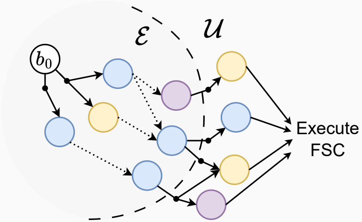
Using FSCs as cut-offs to obtain a better approximation of the unexplored parts of the belief space



Already very non-optimal FSCs improve bounds provided by existing cut-offs techniques

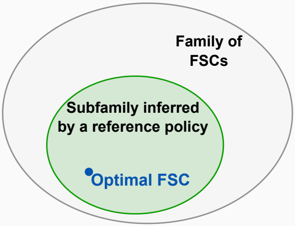
Two Main Ideas of Our Approach

Using FSCs as cut-offs to obtain a better approximation of the unexplored parts of the belief space



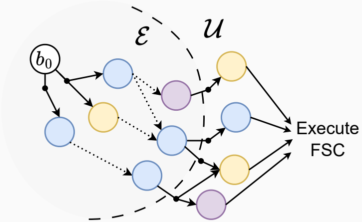
Already very non-optimal FSCs improve bounds provided by existing cut-offs techniques

Using reference policies from belief-space exploration to guide the inductive synthesis search



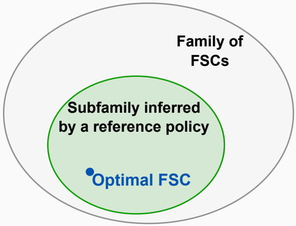
Two Main Ideas of Our Approach

Using FSCs as cut-offs to obtain a better approximation of the unexplored parts of the belief space



Already very non-optimal FSCs improve bounds provided by existing cut-offs techniques

Using reference policies from belief-space exploration to guide the inductive synthesis search

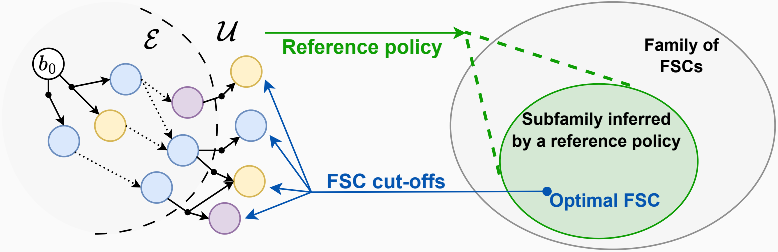


Already a very shallow exploration of the belief-space is useful for prioritising the family exploration.

SAYNT - Novel Symbiotic Synthesis Algorithm

SAYNT is an iterative anytime synthesis algorithm

- closed loop integration of the inductive synthesis and the belief-space exploration
  - PAYNT provides cut-off FSCs for Storm,
  - Storm provides reference policies for PAYNT and suggest where to add the memory
- in each iteration two FSCs $F_{\mathcal{I}}$ and $F_{\mathcal{B}}$ are obtained



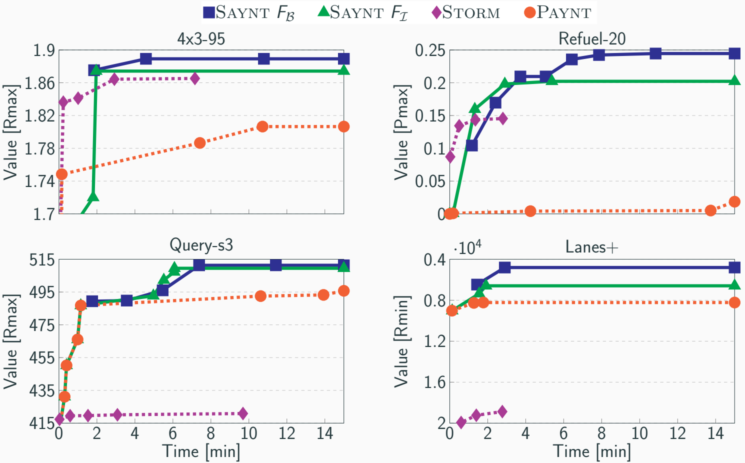
Wide range of benchmarks from AI and formal verification communities

Model	$\|S\|$	$\sum Act$	$\|Z\|$	Spec.	Over-approx.	Model	$\|S\|$	$\sum Act$	$\|Z\|$	Spec.	Over-approx.
4x3-95	22	82	9	$R_{\max}$	≤ 2.24	Drone-4-2	1226	2954	761	$P_{\max}$	≤ 0.98
4x5x2-95	79	310	7	$R_{\max}$	≤ 3.26	Drone-8-2	13k	32k	3195	$P_{\max}$	≤ 0.99
Hallway	61	301	23	$R_{\min}$	≥ 11.5	Lanes+	2741	5285	11	$R_{\min}$	≥ 4805
Milos-97	165	980	11	$R_{\max}$	≤ 80	Netw-3-8-20	17k	30k	2205	$R_{\min}$	≥ 4.31
Network	19	70	5	$R_{\max}$	≤ 359	Refuel-06	208	565	50	$P_{\max}$	≤ 0.78
Query-s3	108	320	6	$R_{\max}$	≤ 600	Refuel-20	6834	25k	174	$P_{\max}$	≤ 0.99
Tiger-95	14	50	7	$R_{\max}$	≤ 159	Rocks-12	6553	32k	1645	$R_{\min}$	≥ 17.8

Implemented in **PAYNT** `https://github.com/randriu/synthesis`

Experimental Evaluation

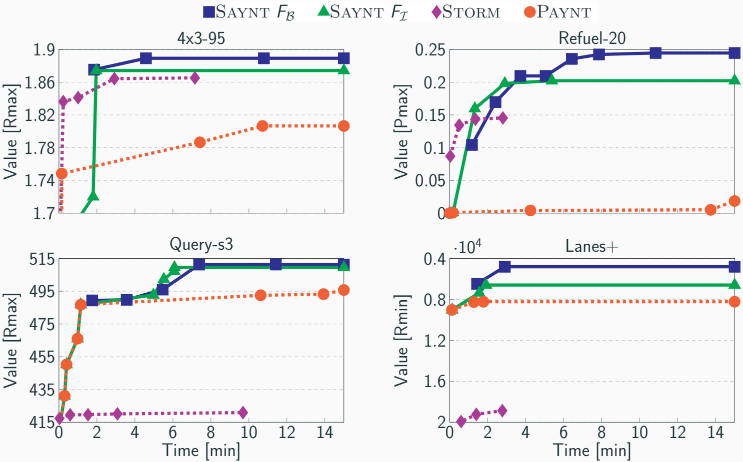
SAYNT vs. state-of-the-art tools (STORM and PAYNT)



SAYNT steadily outperforms both baselines

Experimental Evaluation

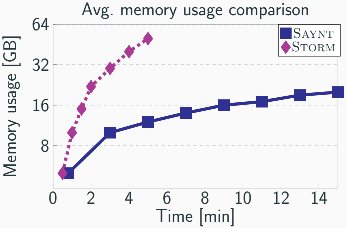
SAYNT vs. state-of-the-art tools (STORM and PAYNT)



The quality of improvements grows with the complexity of POMDPs and reaches up to 40%

Experimental Evaluation

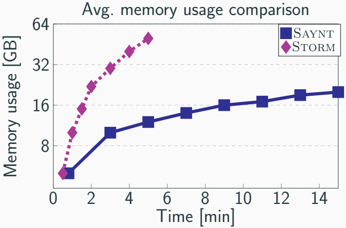
Memory footprint



- SAYNT significantly reduces the memory usage of STORM
- This allows an efficient belief-space exploration of larger POMDPs

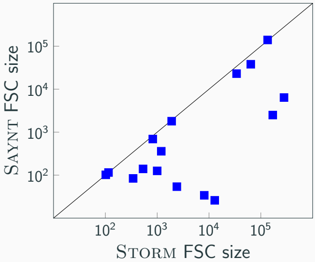
Experimental Evaluation

Memory footprint



- SAYNT significantly reduces the memory usage of STORM
- This allows an efficient belief-space exploration of larger POMDPs

FSC size comparison



- SAYNT produces more compact FSCs compared to STORM while achieving better values

FSC size refers to the encoding size.

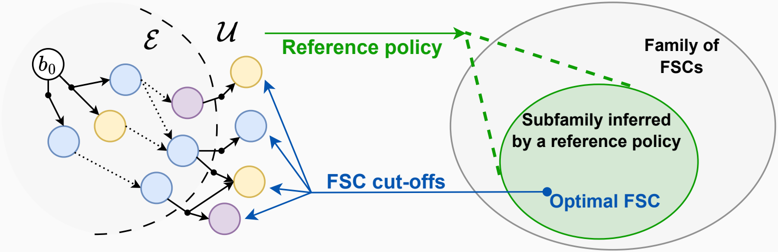
Future research

1. Discounted vs. undiscounted specifications
    - How to properly compare the results for different specifications?
    - Can we use some of the algorithms built for discounted specifications to help our approach and vice versa?
2. Dec-POMDPs and partially-observable stochastic games
    - How to make our framework efficient in more difficult domains?
3. Combination with reinforcement learning approaches
    - Can we use results from RL to help formal methods scale better?
    - Can we formally guarantee the correctness of NN strategies?

Conclusions

Novel algorithm for POMDPs with indefinite-horizon specifications

- symbiotically integrates the belief-space exploration and the inductive synthesis
- outperforms state-of-the-art methods on a wide range of benchmarks
- strengthens the position of formal methods for the POMDP synthesis problem



See: **Andriushchenko et al. Search and Explore: Symbiotic Policy Synthesis in POMDPs. In CAV'23.**