Search and Explore:
Symbiotic Policy Synthesis in POMDPs

Bc. Filip Macák (xmacak07@fit.vutbr.cz)

Supervisor: Assoc. Prof. Milan Češka

Excel@FIT 2023

BRNO FACULTY
UNIVERSITY OF INFORMATION
OF TECHNOLOGY TECHNOLOGY

Partially-observable Markov decision processes (POMDPs)

- important model for sequential decision-making under uncertainty and limited observability
- states with the same observation are indistinguishable

observations:

- exit
- crashed
- near obstacle
- other
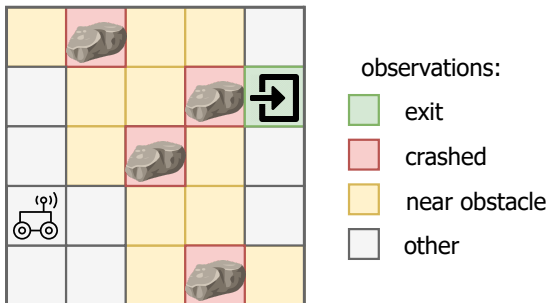
Specification:

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

Find the optimal strategy for the given specification

- indefinite-horizon specifications - no discounts, long-term planning – finding an optimal policy is undecidable
- focus on compact, verifiable and easy-to-use strategies

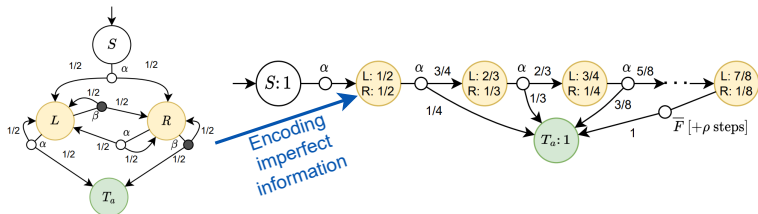**Finite-state controller** (FSC) based on Mealy machines
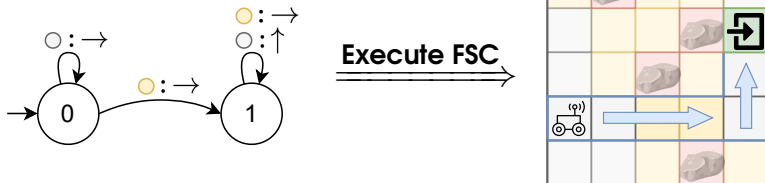
Many practical applications:

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

Belief - probability distribution over the states of a POMDP

Construct and analyse the reachable belief space

- it might be huge or even infinite
- various approximations of the unexplored belief space, namely, cut-offs and point-based

Limitations:

- cut-offs (implemented in the tool Storm) are not sufficient
- point-based methods, notably SARSOP, perform poorly for long-term planning

Inductive exploration of the family of candidate FSCs

- iterative expansion of the family by adding memory
- uses fully-observable abstraction and counter-examples to steer the search
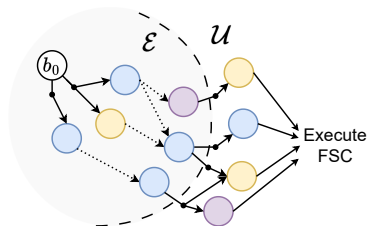- implemented in the tool PAYNT (developed at FIT)
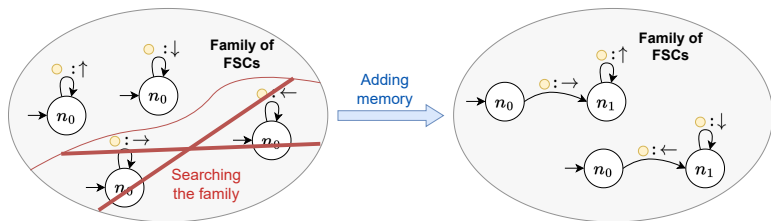
Limitations:

- the family size grows exponentially with the memory (states) added to the FSCs
- if a lot of memory is needed or the POMDP is too large, exploration becomes computationally intractable

Two novel ideas

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

Using reference POMDP policies to guide the inductive synthesis search
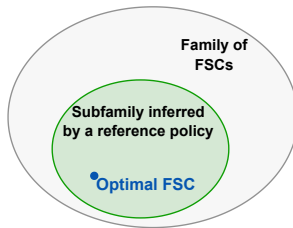
SAYNT is an iterative anytime synthesis algorithm

- combining the inductive synthesis and belief exploration
- closing the integration loop between Storm and PAYNT
- Storm provides reference policies for PAYNT, and PAYNT provides cut-off FSCs for Storm
- in each iteration two FSCs $F_{\mathcal{I}}$ and $F_{\mathcal{B}}$ are obtained
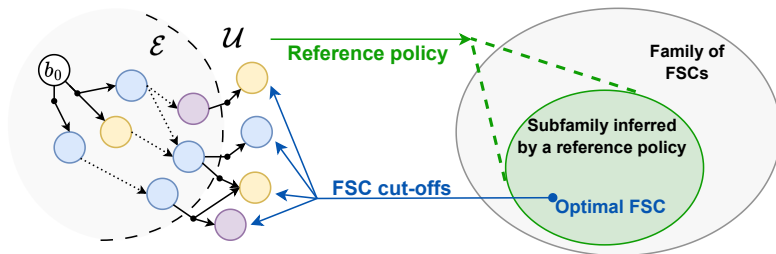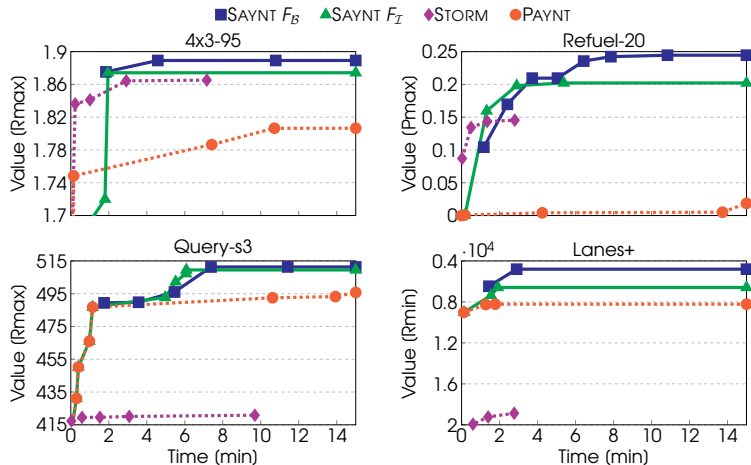
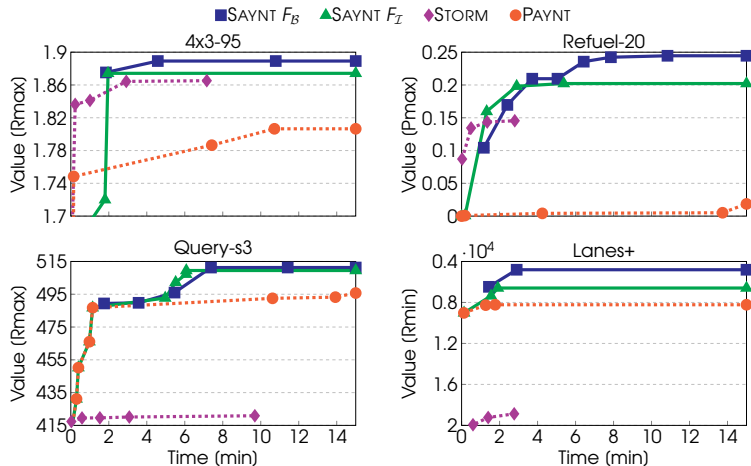## SAYNT vs. state-of-the-art tools (Storm and PAYNT)
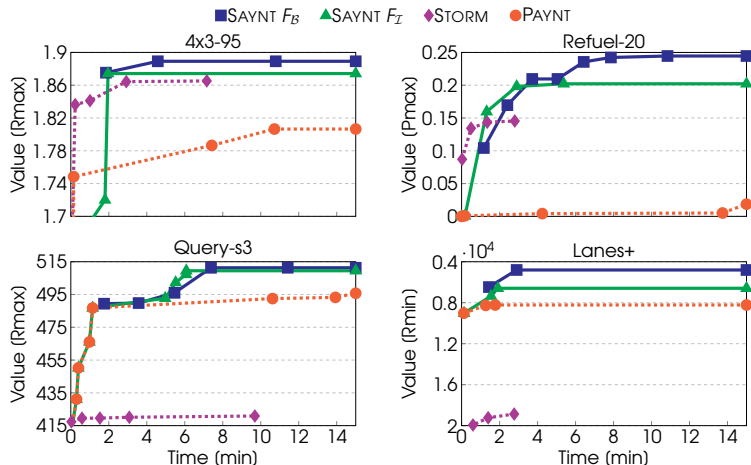
SAYNT steadily outperforms both baselines on a wide range of benchmarks from AI and formal verification communities

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%

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

SAYNT gives the users a unique choice of which FSC to use, smaller $F_{\mathcal{I}}$ or slightly better but much larger $F_{\mathcal{B}}$

This work has been accepted to CAV'23 (A* conference)

I would like to thank all the co-authors:
- Milan Češka   (**BUT FIT**)
- Roman Andriushchenko   (**BUT FIT**)
- Alexander Bork   (RWTH Aachen University)
- Sebastian Junges   (Radboud University)
- Joost-Pieter Katoen   (RWTH Aachen University)

Conclusions

We present a novel symbiotic algorithm SAYNT that is built on two novel ideas

- SAYNT outperforms current state-of-the-art methods on a wide range of benchmarks
- the presented ideas are not tool specific

This work strengthens the position of formal methods for the POMDP synthesis problem

Future research:

- discounted vs. undiscounted specifications
- FSC minimisation
- extracting more information from reference policies