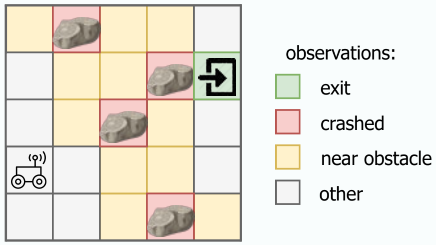
Improving Synthesis of Finite State Controllers for POMDPs Using Belief Space Approximation

Author: Ing. Filip Macák   Supervisor: Assoc. Prof. Milan Češka

BRNO UNIVERSITY OF TECHNOLOGY — FACULTY OF INFORMATION TECHNOLOGY

1. Motivation and Problem Statement

Partially observable Markov decision process (POMDP) is an important model for sequential decision-making under uncertainty and limited observability. These models are widely used in many areas including AI and autonomous planning but have their use even in medicine. Finding good and provably correct solutions is therefore crucial when we consider **safety-critical applications**.



The robot has only a limited observability:

- does not have the access to its location
- only recognises different observations (colours)

Robot's specification:

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

We consider **long-term planning** using indefinite-horizon specifications (i.e. no time bounds and no discounting). Finding the optimal strategy for these specifications is **undecidable** and we focus on finding the best strategy in the given time. We seek for **easy-to-execute and verifiable** strategies and thus encode the strategies as **finite-state controllers (FSCs)** – a variant of Mealy automaton.

2. State of the Art and Its Limitations

Inductive synthesis of FSCs [1] is a formal approach based on a symbolic exploration of a *family of candidate FSCs*. It iteratively expands the family by adding memory to the candidate FSCs.

Limitations:

- for large POMDPs, the family size is huge and its exploration is expensive
- the family size grows exponentially with the memory added to FSC – if a lot of memory is needed, exploration becomes computationally intractable

Belief-based methods, widely used in AI community, explore the belief-space given by probability distributions over states of the POMDP. The belief space might be very large or infinite, therefore various approximations techniques exist, notably, *cut-offs* and *point-based* techniques.

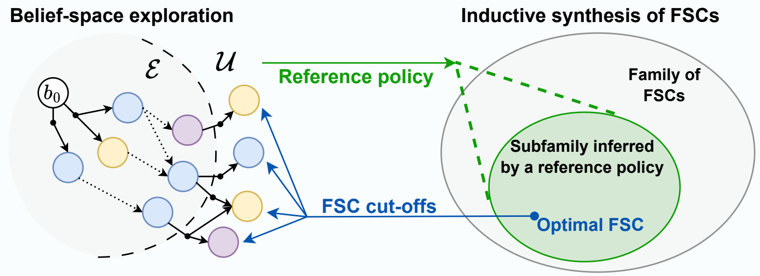
Limitations:

- existing cut-offs (implemented in the tool Storm [2]) require to explore a large belief space
- point-based approximations, notably SARSOP [3], perform poorly for long-term planning

3. Key Contribution: Symbiotic Integration of the State-of-the-art Approaches

The proposed approach builds on the **two novel ideas**:

- use the FSCs obtained from inductive synthesis to improve the cut-offs in the belief-space
- use policies obtained from the explored belief-space to steer and accelerate the inductive search



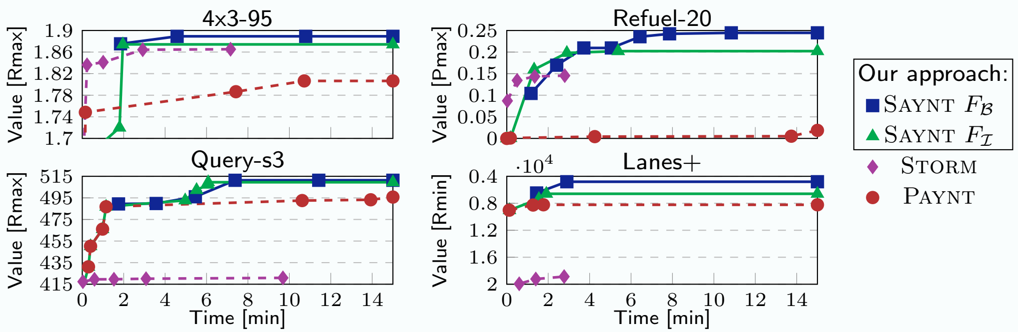
The proposed synthesis algorithm, called Saynt, closes the integration loop:

- the belief-space exploration providing reference policies to help steer the inductive synthesis
- the synthesised FSCs are used to effectively approximate the unexplored belief-space

The algorithm works iteratively, in each iteration two FSCs ($F_{\mathcal{B}}$ and $F_{\mathcal{I}}$) are obtained.

4. Experimental Evaluation

The performance of Saynt is compared to Storm [3] and Paynt [1], state-of-the-art tools for the controller synthesis in POMDPs with indefinite-horizon specifications. **A wide range of benchmark problems** from AI and formal methods communities are used. The graphs show how the quality of the controllers improves over time for selected problems:



Main Result: Our approach steadily outperforms the state-of-the-art tools – the improvements of the final controllers grow with the complexity of POMDPs and reach up to 40%.

Additional key observations:

- Saynt reduces the memory usage of Storm by a factor of 4 and thus allows an efficient belief-space exploration of larger POMDPs (Storm terminates before the time-out with an out of memory signal).
- Saynt produces more compact FSCs compared to Storm while achieving better values.

5. Publication

Based on this thesis, we published the following conference paper:

R. Andriushchenko, A. Bork, M. Češka, S. Junges, J.P. Katoen, and F. Macák. Search and Explore: Symbiotic Policy Synthesis in POMDPs. In CAV'23 (A conference).*

We thank the co-authors for their help (the alphabetical order of the authors is used).

6. References

[1] R. Andriushchenko et al. Inductive synthesis of finite-state controllers for POMDPs. In UAI'22.
[2] A. Bork et al. Under-approximating expected total rewards in POMDPs. In TACAS'22.
[3] H. Kurniawati et al. SARSOP: Efficient point-based POMDP planning by approximating optimally reachable belief spaces. In Robotics: Science and Systems 2008.