Using Reinforcement Learning and Inductive Synthesis for Designing Robust Controllers in POMDPs

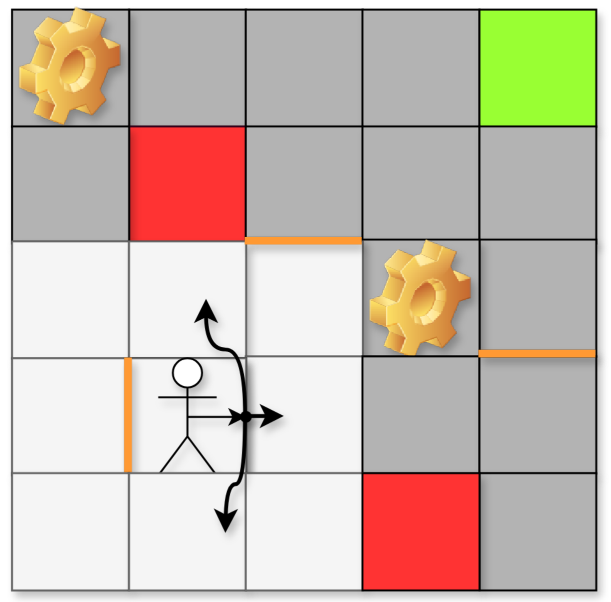
Author: Ing. David Hudák Supervisor: doc. RNDr. Milan Češka, Ph.D.

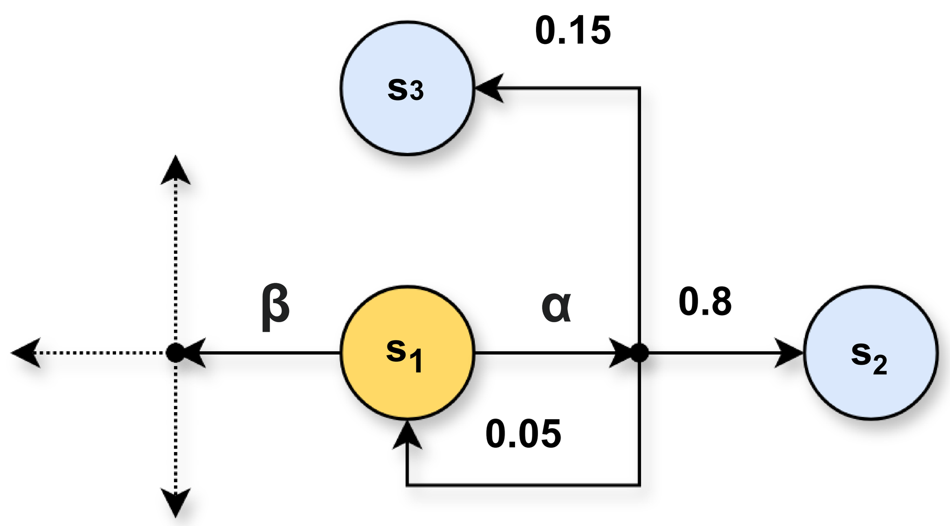
BRNO UNIVERSITY OF TECHNOLOGY
FACULTY OF INFORMATION TECHNOLOGY

Problem Introduction

Partially Observable Markov Decision Process (POMDP) serves as a formal framework for depicting models engaged in sequential decision-making tasks under uncertainty with limited observations. These models encompass challenging fields like autonomous driving, healthcare, video-games, financial markets, and many others, where our observation from the environment is limited.

In environments with uncertainty, controllers depend on state estimation derived from prior observations (frames, symptoms, sensor data, etc.). We can estimate it by computing the **belief** based on a known model, updating the state of a **finite state controller (FSC)**, or approximating it by a **recurrent neural network (LSTM)** trained with reinforcement learning algorithms.





Determining the best strategy for POMDPs is generally undecidable, and in our thesis, we focused on the improvement of two very different state-of-the-art approaches through their integration.

Current Approaches and Challenges

Model-based formal methods including **inductive synthesis of FSCs** or **belief-based** approaches operate under the assumption that the model is known. The FSC synthesis method ihard FSCnvolves exploring families of candidate FSCs to provide reliable and verifiable observation-based controllers for tasks with uncertainty with simple inference process. However, it composes several limitations:

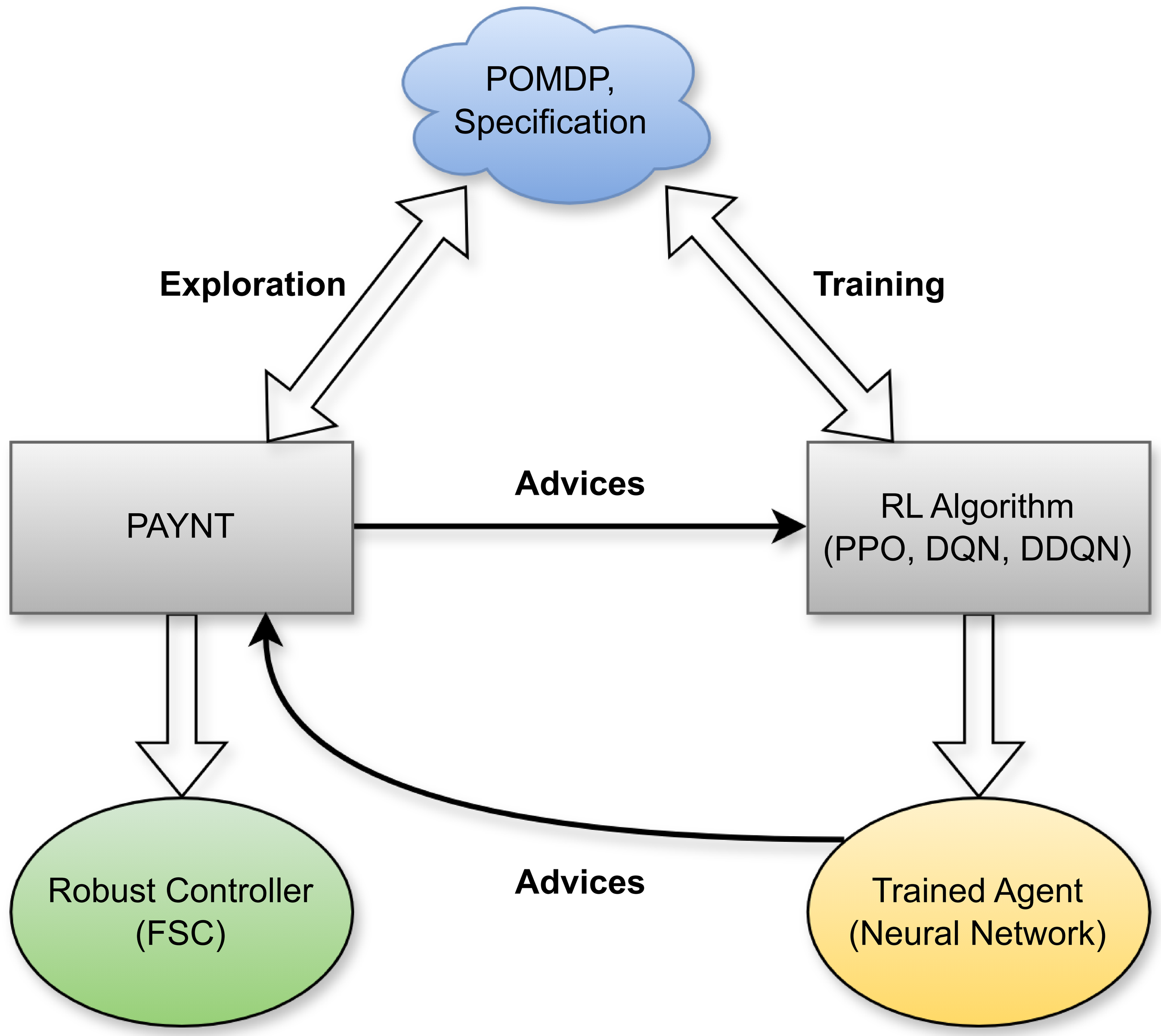
- **Exponentially-Large Design Space**: Exploring the design space is difficult.
- **Scalability**: At least one node for each observation.
- **Memory Estimation**: Size of the optimal FSC is unknown.

(Deep) Reinforcement learning is an approach based on training agents by performing actions in environments and learning from the feedback (rewards) obtained. State-of-the-art approaches for POMDPs are based on recurrent neural networks (LSTM, GRU) combined with algorithms such as **PPO**, **(D)DQN** or **SAC**. However, current implementations face many challenges:

- **Sparse Reward**: Environments offer rewards only after achieving the goal state, which can be challenging to discover with an initial random policy.
- **Stability and Reproducibility**: The learning process of current SOTA algorithms is usually unstable and the results are hard to reproduce.
- **Explainability**: Neural networks are hard to explain/verify.

Proposed Approach

We introduce an innovative method that merges model-based inductive FSC synthesis with model-free reinforcement learning to achieve optimal policies regarding scalability and verifiability. The implementation consists of two distinct learning processes, the first based on **FSC exploration**, and the second based on **deep reinforcement learning**.



The implemented solution operates in two distinct **modes**:

One-Time Advice: Our objective is to train the optimal agent using reinforcement learning techniques and then offer hints to PAYNT FSC synthesis through **pruning of the FSC design space**. We aim to extract the best verifiable policy from the agent. This explored domain is in modern research called **surrogate** solutions.

Closed-Loop: Agents trained with reinforcement learning provide hints to PAYNT, and PAYNT provides hints (sampled trajectories) to agents. Our goal is to iteratively improve both approaches.

Experimental Evaluation

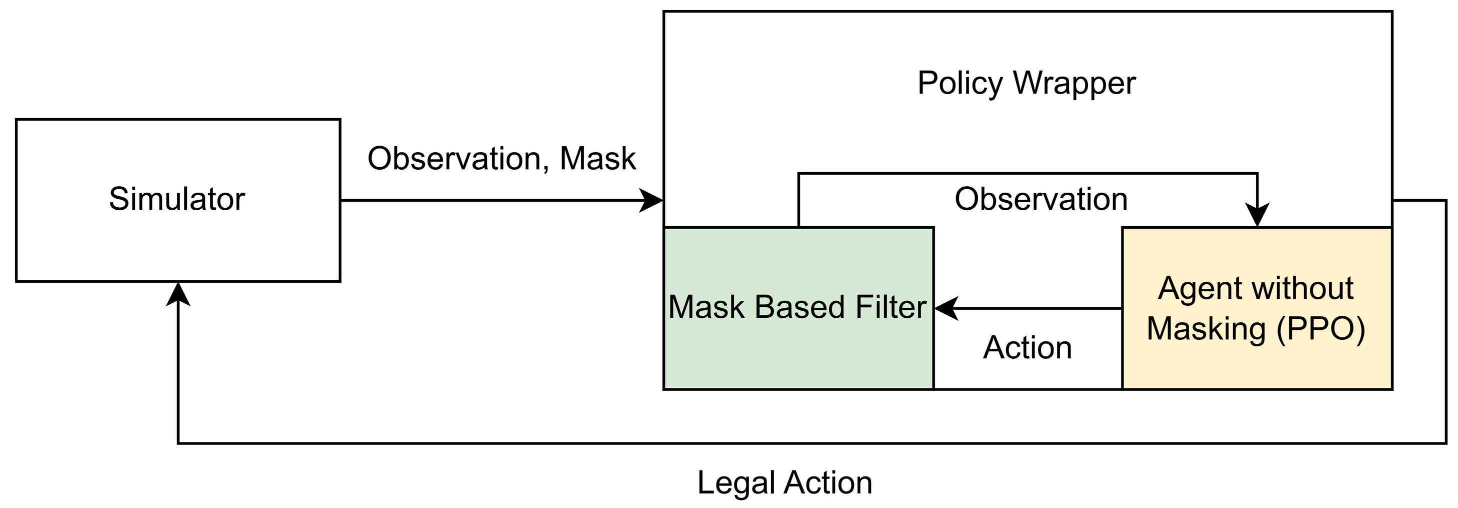
Benchmarks: Performed experiments with various models of grid environments with different levels of observability, complexity and size focused on long-term planning. We also evaluated our implementation on complex network model.

State-of-the-Art RL Results: In the provided benchmarks, our one-time advice, closed-loop, and independent RL algorithms surpassed the performance of existing implementations.

Novel Encoding Method: We proved that using different encoding methods can significantly improve the training process for formally defined models.

Technical Details and Our Contributions

- Implementation of a new **RL toolkit** using TensorFlow Agents.
- Implementation of an **environment wrapper** using the Storm model and simulator representation.
- Novel method for **interpretation of neural agents** through approximations from multiple trajectory observations of the trained agent.
- Introduction of a novel approach for **soft** and **hard FSC** hints in reinforcement learning through FSC-based policy sampling.
- Investigation of various **encoding strategies** for reinforcement learning, including basic integer, one-hot encoding, and encoding based on **valuations** from the Storm simulator.
- Implementation of a novel **policy-wrapping** combined with *masking* technique to handle dynamic action space to reduce the exploration space of the reinforcement learning agents.



Future Work

Publication: Currently working on a publication based on results from diploma thesis and an improved version of PAYNT – SAYNT.

Network Interpretation: Implementation of more complex policy extraction method from neural network, such as quantized bottleneck extraction, or adjusting the network architecture with soft-max layers to improve explainability.

Imitation Learning: Extension of current techniques of hints from model-based PAYNT approach to further improve performance of reinforcement learning agents.

References

[1] Roman Andriushchenko, Bork Alexander, Milan Češka, Sebastian Junges, Joost-Pieter Katoen, and Filip Macák.
Search and explore: Symbiotic policy synthesis in pomdps.
In *Computer Aided Verification*, volume 13966 of *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, pages 113–135. Springer Verlag, 2023.

[2] Steven Carr, Nils Jansen, Sebastian Junges, and Ufuk Topcu.
Safe reinforcement learning via shielding under partial observability.
In *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence*, AAAI'23/IAAI'23/EAAI'23. AAAI Press, 2023.