

Università degli Studi di Verona
Master Degree in Scienze e Ingegneria Informatica

A Rule-Based Approach for Safe Information- Oriented Planning.

Date:15/03/2022

Presented by
Nereo Sorio

Supervisor **Prof. Alessandro Farinelli**
Co-Supervisor **Giulio Mazzi**

CONTENTS

1. Motivation
2. Basic Knowledge
3. Thesis Contribution
4. Evaluation
5. Conclusions & Future Works

MOTIVATION



Mars Rover



Autonomous Self Driving Car

Planning is crucial to devise a sequence of actions to achieve a given goal.

We focus on a specific problem:

sequential decision making under uncertainty (POMDPs / ρ -POMDPs).

BASIC KNOWLEDGE

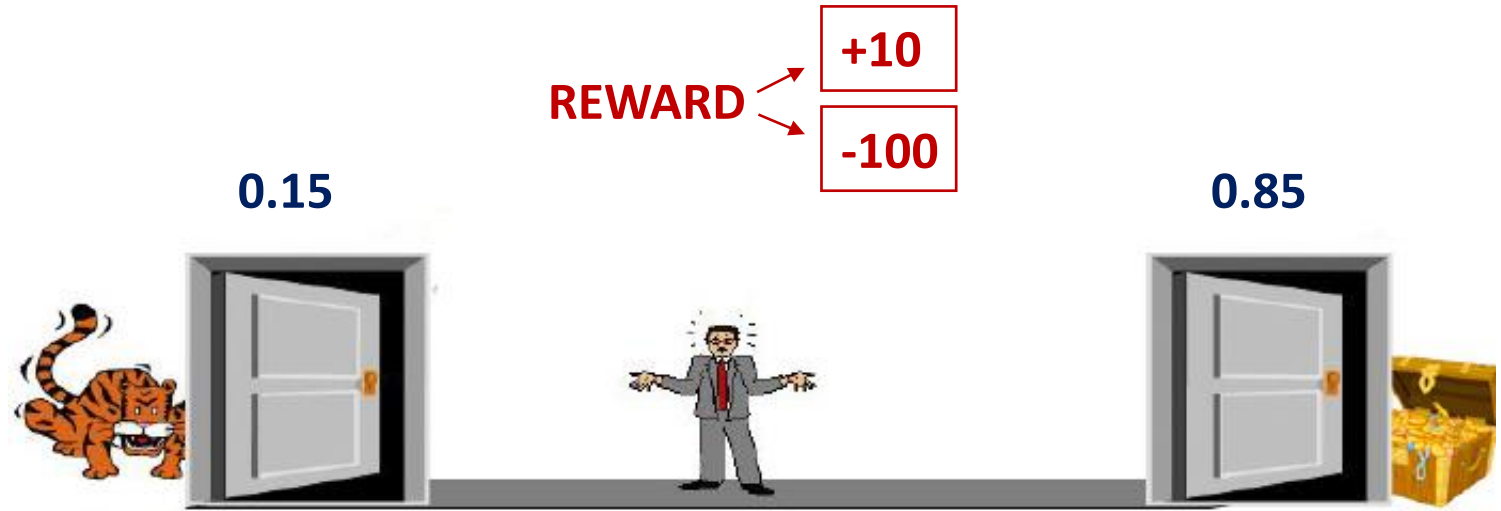
Overview of POMDPs

Extension of **Markov Decision Processes** that handles partial observability.

ACTIONS:

- OPEN-LEFT
- OPEN-RIGHT
- LISTEN

The uncertainty influences the reward indirectly.



SEQUENCE 1

- Listen
- Open – Right

REWARD = +9 / -101

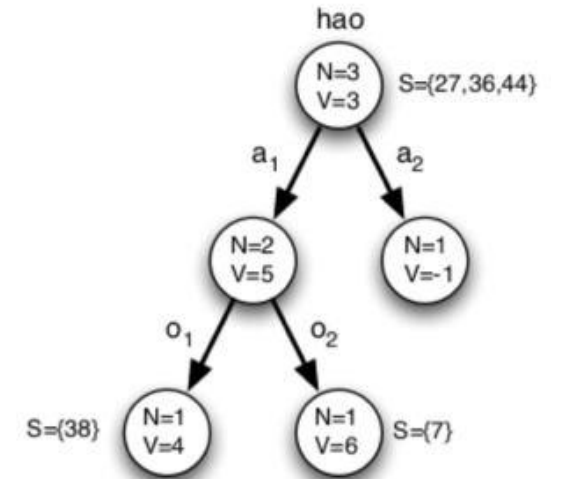
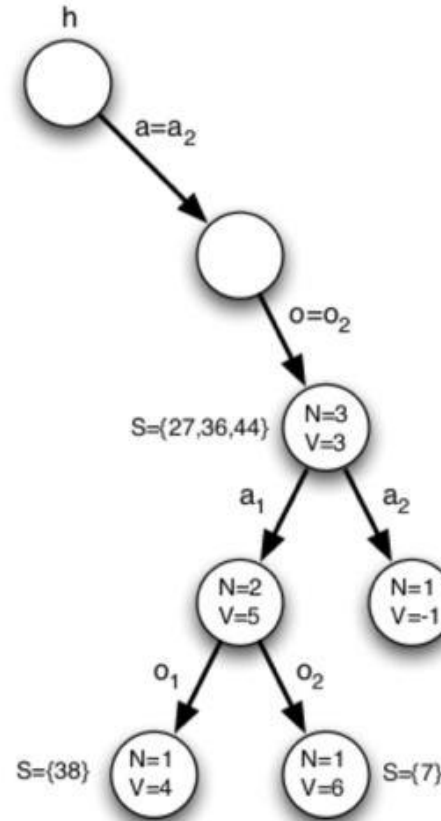
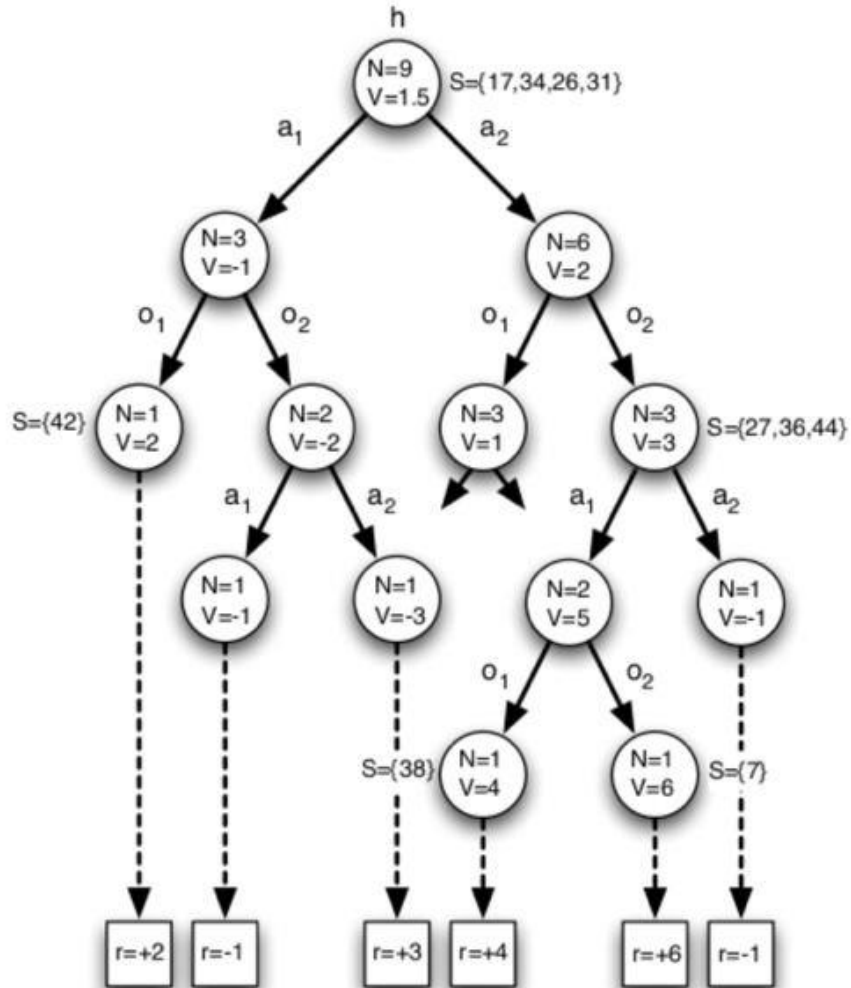
SEQUENCE 2

- Listen
- Listen
- Listen
- Open – Right

REWARD = +7 / -103

BASIC KNOWLEDGE

Overview of POMCP

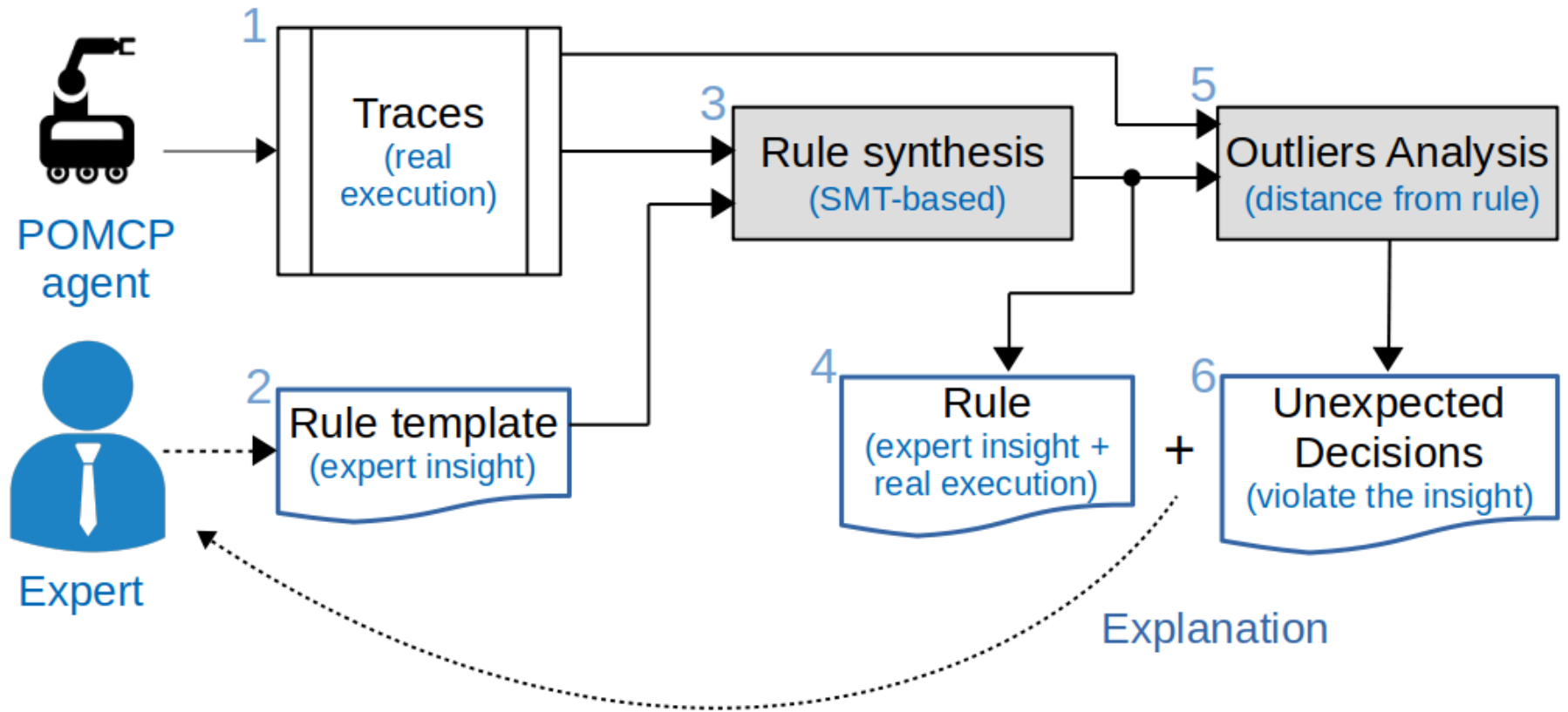


- Monte – Carlo based.
- It generates approximate and online policies.
- It scales to large instances.
- Hard to explain.

BASIC KNOWLEDGE

Overview of Explainable POMCP (XPOMXP)

- Extract rules that give a **compact representation** of the POMCP policy.
- It uses traces that collect **belief action** pairs.



BASIC KNOWLEDGE

Overview of Explainable POMCP (XPOMXP)

- The goal is to define the **missing elements** (the confidence level of the door we want to open).
- The algorithm automatically find them **observing the traces**.

RULE TEMPLATE FORMULA

declare - var x_1, x_2, x_3, x_4 prob;

declare - rule

action Listen $\longleftrightarrow p(\text{right}) < x_1 \wedge p(\text{left}) < x_2$;

action Open_R $\longleftrightarrow p(\text{right}) > x_3$;

action Open_L $\longleftrightarrow p(\text{left}) > x_4$;

where $(x_1 = x_2) \wedge (x_3 = x_4) \wedge (x_3 > 0,9)$.

BASIC KNOWLEDGE

Overview of ρ -POMDPs

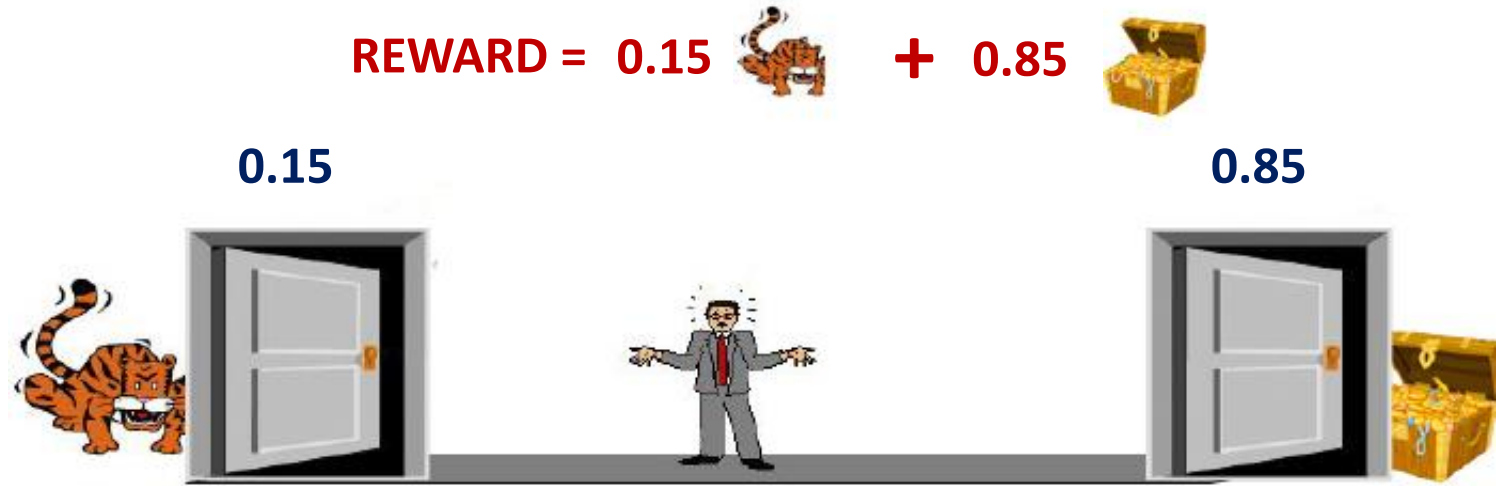
Extension of POMDPs that directly quantifies the influence of the belief (**belief dependent**).

Allow to define not only **control oriented objectives**, but also **information-oriented** ones.

A recent work for solve ρ -POMDPs is **ρ -POMCP** which is an **extension of POMCP**.

BASIC KNOWLEDGE

Overview of ρ -POMCPs



ACTIONS:

- OPEN-LEFT
- OPEN-RIGHT
- LISTEN

SEQUENCE 1

- Listen
- Open – Right

REWARD =



SEQUENCE 2

- Listen
- Listen
- Listen
- Open – Right

REWARD =



THESIS CONTRIBUTION

1.

Extending
POMCP to
Handle
 **ρ -POMDPs &
Rule Synthesis.**

2.

**Application
Domains.**

3.

Generation of
**Extended
Traces.**

THESIS CONTRIBUTION

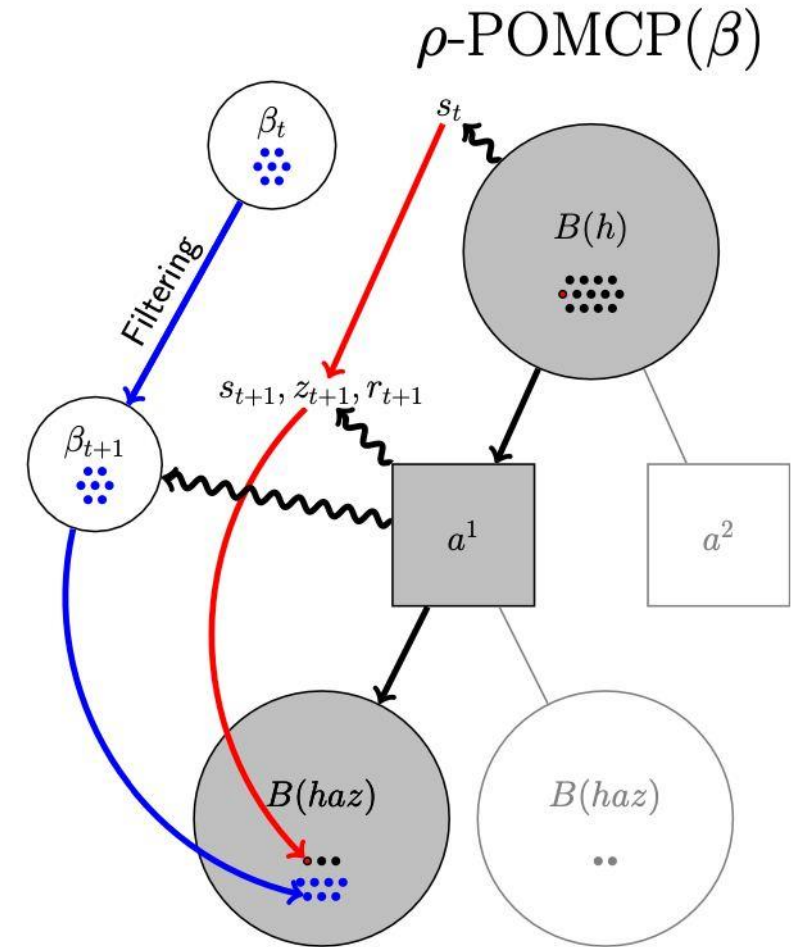
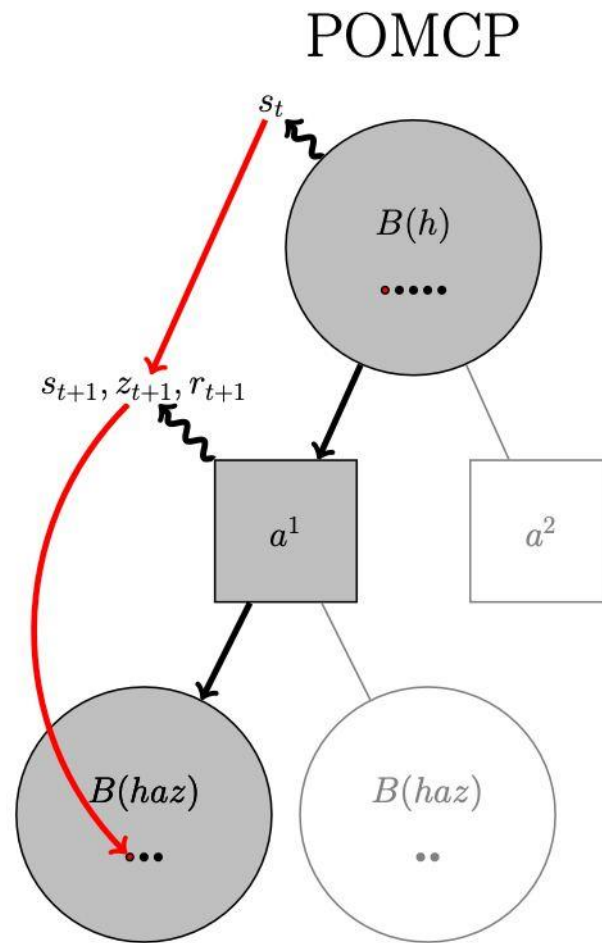
1. Extending POMCP to Handle ρ -POMDPs

BELIEF



BAG

Implementation
available on GitHub.



THESIS CONTRIBUTION

1. Extending POMCP to Handle ρ -POMDPs

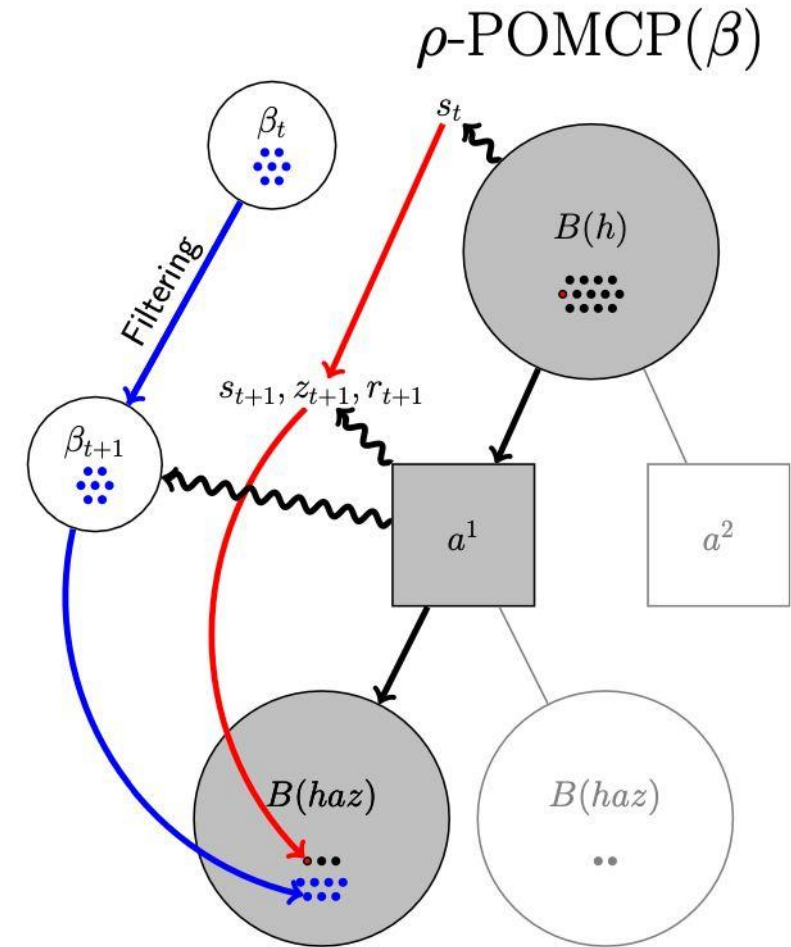
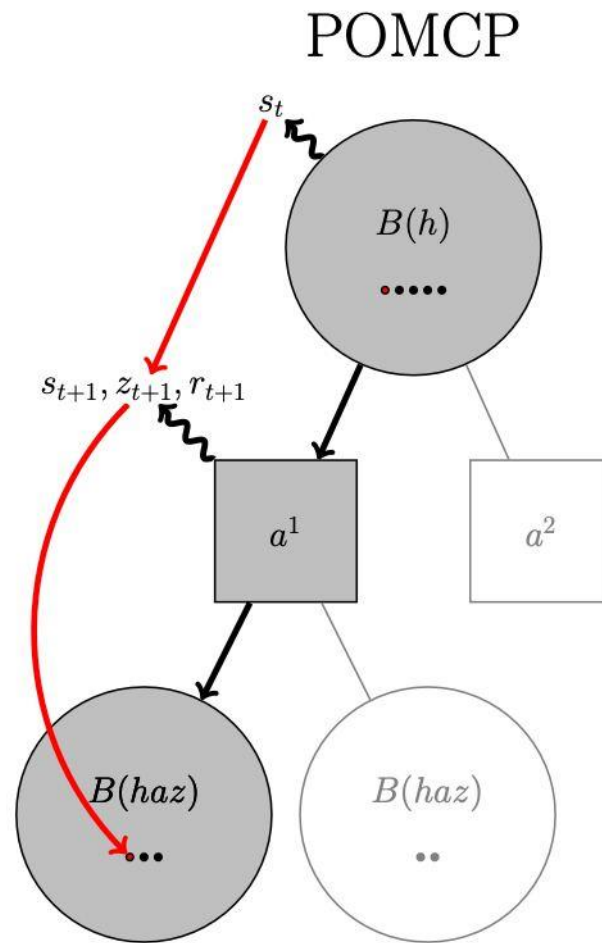
BELIEF



BAG

+ MORE FLEXIBLE
+ MORE EXTENSIVE
+ MORE ESTIMATES

Implementation
available on GitHub.



THESIS CONTRIBUTION

2. Application Domains

ρ - TIGER

REWARD = 0.85  + 0.15 



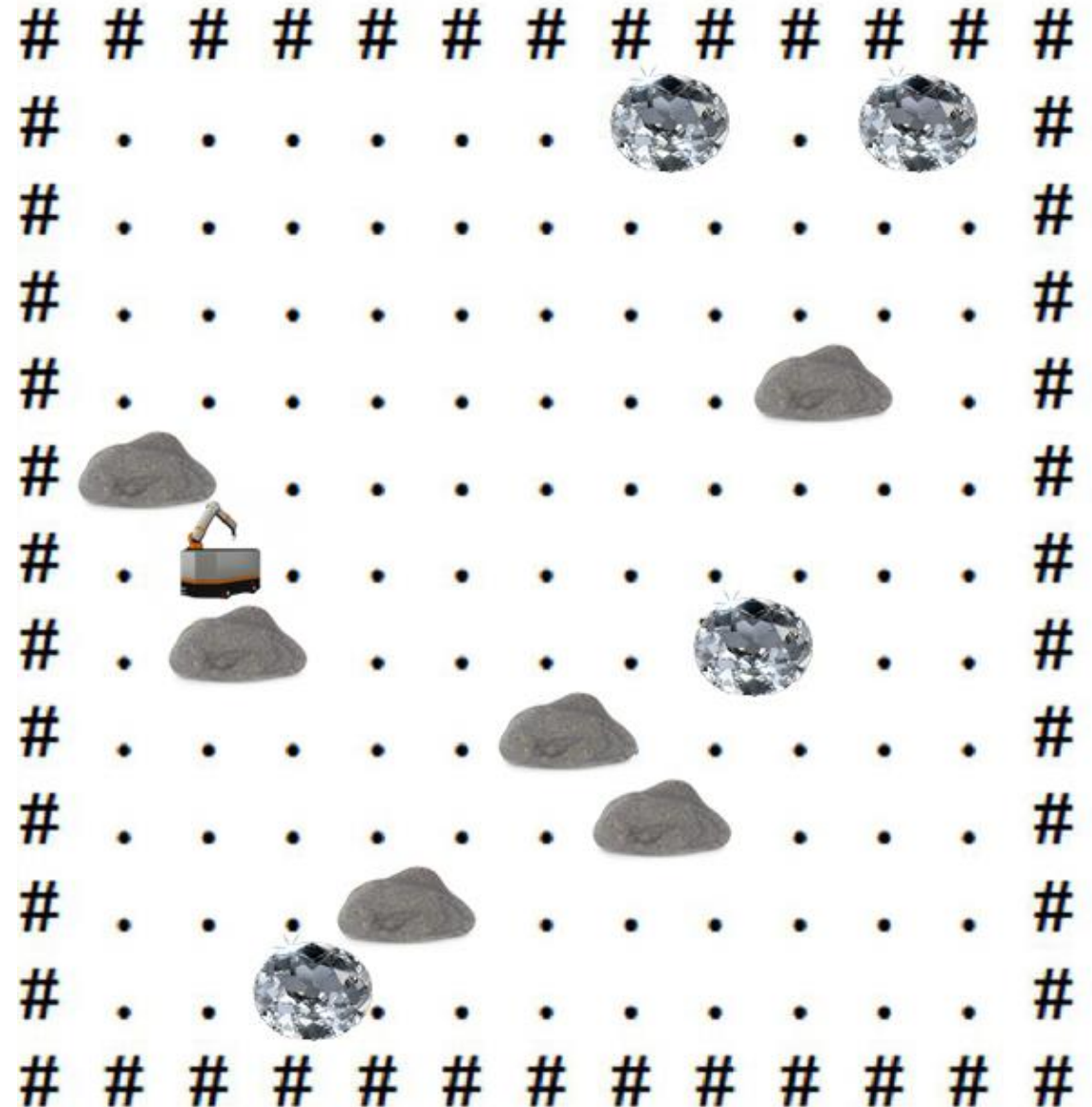
THESIS CONTRIBUTION

2. Application Domains

ρ - ROCKSAMPLE

ACTIONS:

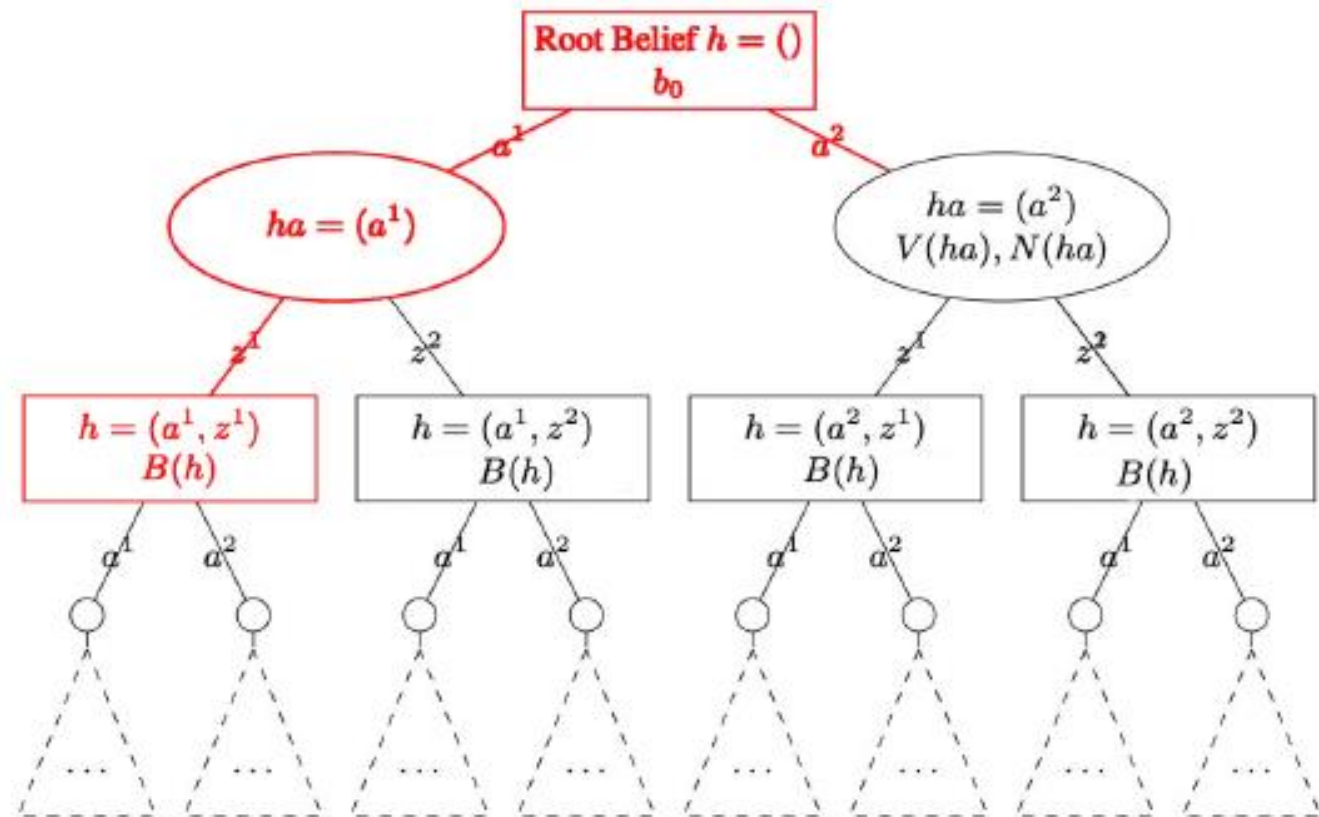
- NORTH
- SOUTH
- WEST
- EAST
- CHECK_ROCK_I
- SAMPLE_ROCK



THESIS CONTRIBUTION

3. Generation of Extended Traces with STANDARD TRACES

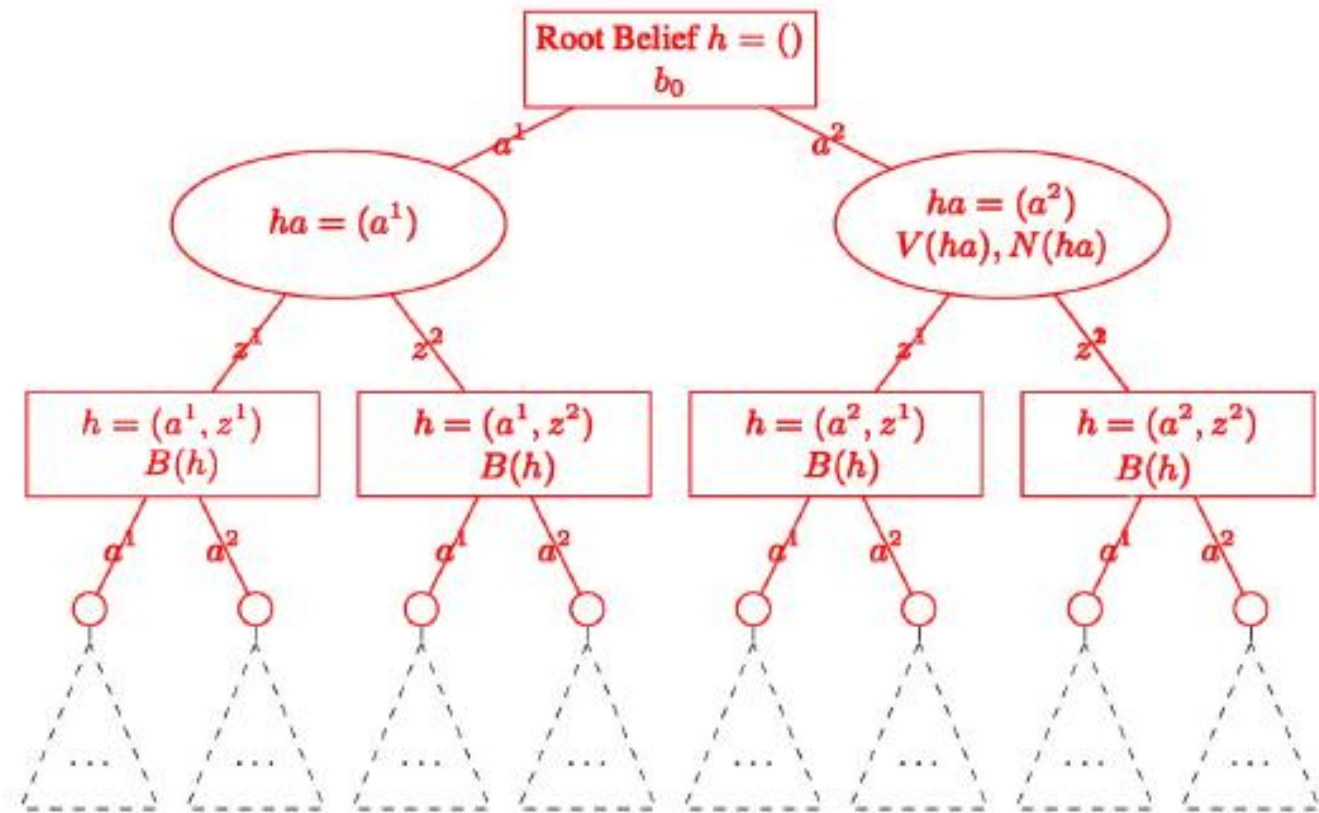
- Belief of the **Root Node**
- Action taken
- Perceived Observation
- Reward Obtained



THESIS CONTRIBUTION

3. Generation of Extended Traces with STANDARD TRACES

- Belief of the **Next Node**
- Action that **will be** taken
- Observation that **will be** perceive
- Reward that **will be** obtain
- List of values of all possible actions starting from the next node



EVALUATION

- ρ – TIGER
- ρ - ROCKSAMPLE

EVALUATION

ρ - TIGER

c	x_1	x_3	n_E	total
20	0.762881	0.918019	53	214
50	0.825119	0.958646	32	475
80	0.95794	0.962695	1	525
110	0.955195	0.955373	0	514

RESULT AFTER RULE SYNTHESIS.

c	Discounted return	Undiscounted Return
20	-8.53	-8.74
50	2.02	2.95
80	4.22	5.75
110	3.24	4.56

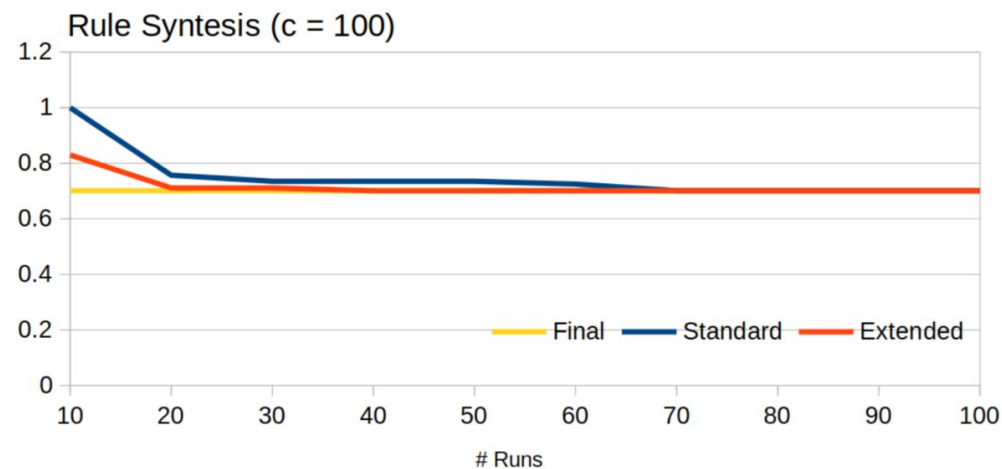
REWARDS.

EVALUATION

ρ -ROCKSAMPLE

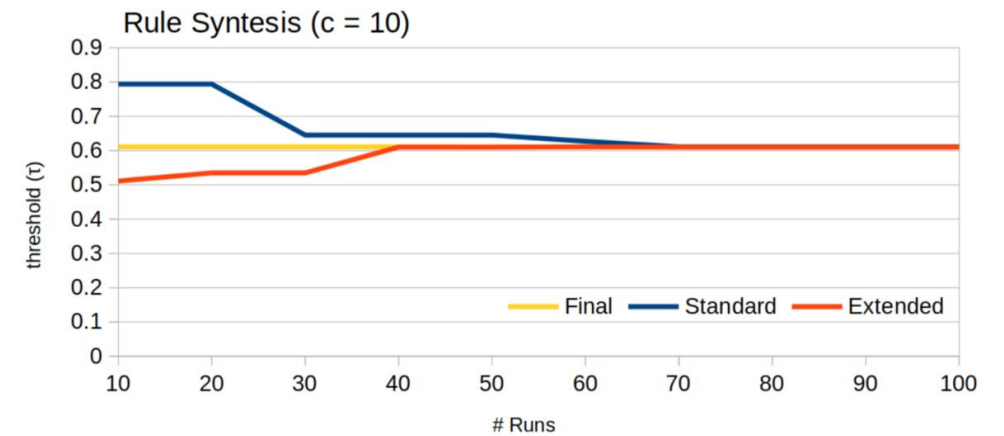
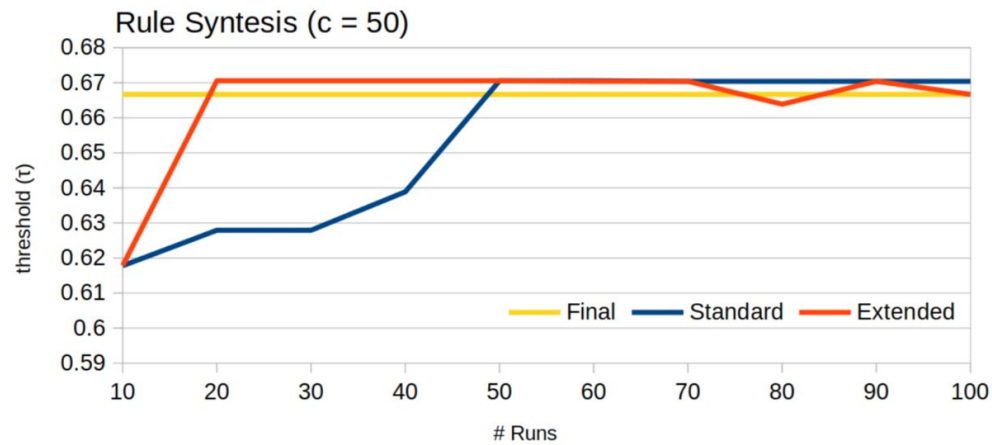
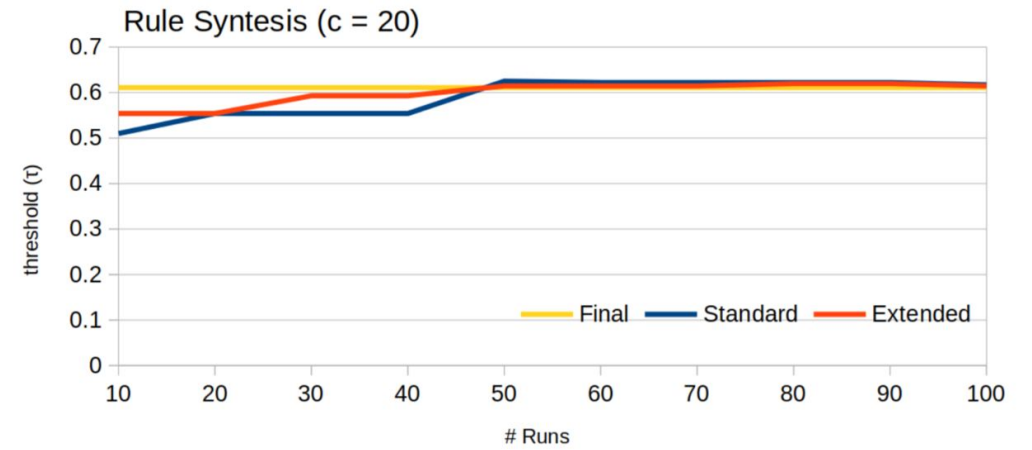
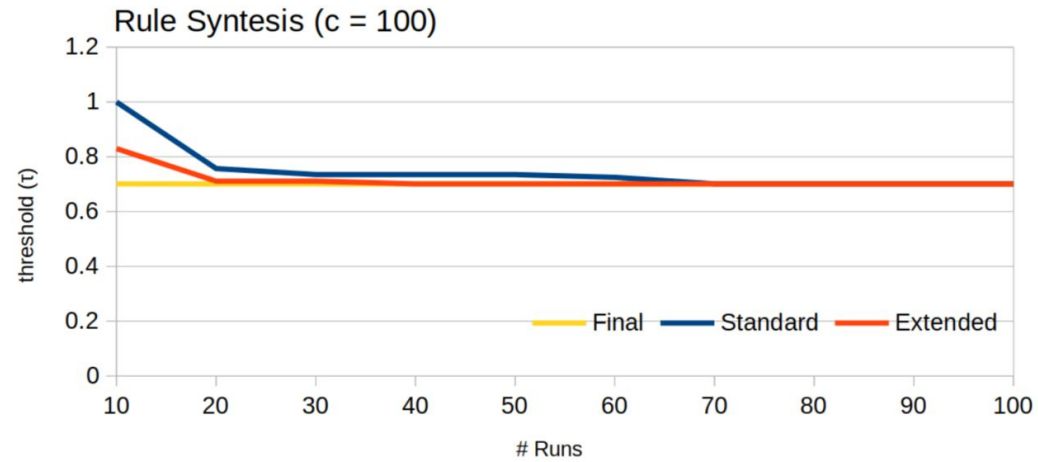
Run	Standard Traces			Extended Traces		
	τ	n_E	total	τ	n_E	total
10	1	10	122	0.829703	15	140
20	0.757	14	256	0.711	27	291
30	0.735	21	399	0.711	39	450
40	0.735	26	528	0.701	49	594
50	0.735	34	631	0.701	61	709
60	0.725	43	782	0.701	81	884
70	0.701	51	915	0.701	95	1038
80	0.701	61	1059	0.701	106	1197
90	0.701	72	1203	0.701	120	1352
100	0.701	82	1381	0.701	133	1546

→ **REWARD RANGE = 100**



EVALUATION

ρ -ROCKSAMPLE



CONCLUSION

This Integration between ρ -POMCP algorithm and XPOMCP is particularly relevant in complex problems, such as rock-sample, which contain thousands of possible states.

Thanks to the new method we can build rules that explain the behaviour the policy using fewer runs than POMCP.

Rule generation can be useful to better understand ρ -POMCP generated policies, and it is possible to use the properties to improve the process of trace generation.

FUTURE WORKS

1.

Apply the extended framework to **other domains.**

2.

Defines rules that specify **temporal properties.**

3.

Implement extended traces in **classical POMCP.**

THANK YOU
FOR YOUR
ATTENTION.