@ ICML | 2022

Leveraging Approximate Symbolic Models for Reinforcement Learning via Skill Diversity

Lin Guan*, Sarath Sreedharan*, Subbarao Kambhampati
School of Computing & AI, Arizona State University
lguan9@asu.edu

Integrating Symbolic Planning and RL

Symbolic Planning Based Methods

Examples: PDDL, STRIPS
Pros:
- A natural way to express human knowledge about actions

Cons:
- Hard to capture all the details of the task and environment

Reinforcement Learning Based Methods

Examples: DQN, TRPO, SAC
Pros:
- Can start from scratch

Cons:
- Extremely high sample complexity

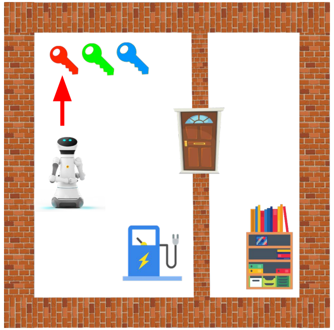
Integrating Symbolic Planning & RL

- Guide RL with symbolic knowledge/advice (better sample efficiency)
- A natural interface for humans to specify goals & constraints (i.e. to define task rewards)

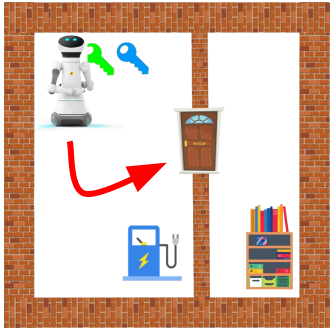
Integrating Symbolic Planning and RL

Learn temporally extended operators (**options**) for symbolic actions:

Pickup Key:
Precondition(s): N/A
Effect(s): has-key

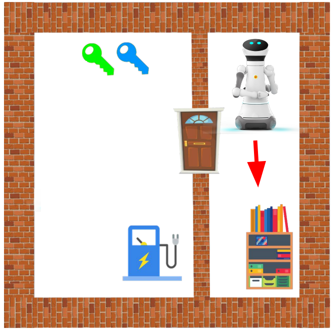


Open Door:
Precondition(s): has-key
Effect(s): door-open



...

Go to Shelf:
Precondition(s): at-right-room
Effect(s): at-destination



Execute the learned options sequentially according to the symbolic plan

Integrating Symbolic Planning and RL

Learn temporally extended operators (**options**) for symbolic actions:

Pickup Key:
Precondition(s): N/A
Effect(s): has-key

Open Door:
Precondition(s): has-key
Effect(s): door-open

...

Go to Shelf:
Precondition(s): at-right-room
Effect(s): at-destination

Most prior works assume the symbolic models are **correct and complete**

But we can't guarantee this in practice!

Sources of Incorrectness:

- Human mistakes
- Plans/models given by other ML models, e.g., LLMs

Example 1: Partially Specified Precondition(s)

- The human may overlook the fact that only the blue key can open the door:

```
(:action open_door
         :parameters ()
         :precondition (has-key)
         :effect (and (door-open)))
```

- If the robot myopically learns a policy to pick up a key, it will only pick up the nearest key (which is the wrong key)

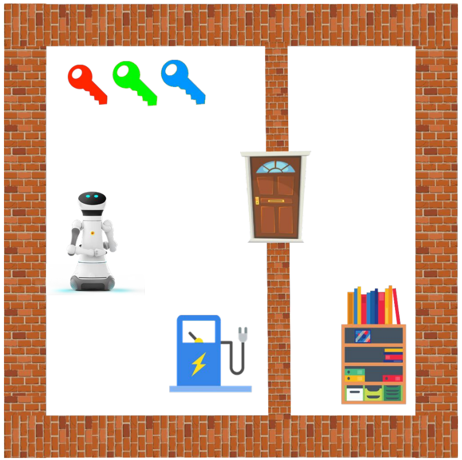


Fig. 1. The Household environment.

Example 2: Incorrect Action Effect(s)

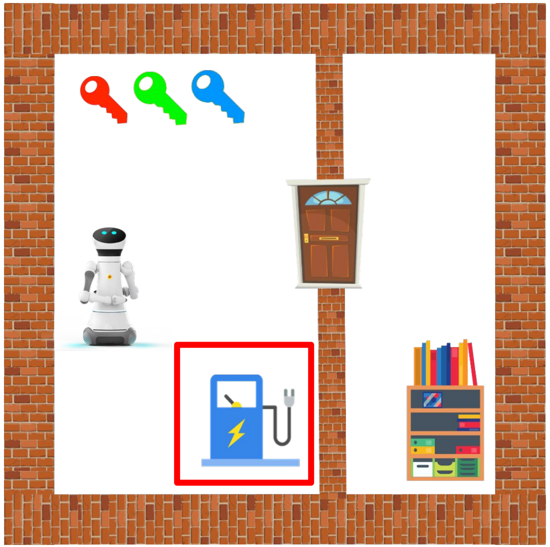
- There may not be one exact state that satisfies all action effects:

```
(:action pass_through_door
        :parameters ()
        :precondition (and )
        :effect (and (at-right-room)
                     (door-ajar)))
```

- But no actual low-level state with `at-right-room` and `at-right-room` being True at the same time, because the door will close once the robot enters the room.
- So no option can be learned for `pass_through_door`.

Example 3: Completely Missing Feature(s)

- The human might not know that this particular robot model has limited battery capacity.
- State variable related to the charging dock is completely missing.



Approximate Symbolic Models Guided RL (ASGRL)

Extracting Task-Hierarchy Information from an Approximate Symbolic Model

- Given the model, we extract **fact landmarks** and their **relative orderings.**
- Landmark information holds in **all plans** for the symbolic model and are thus reliable sources of information about the underlying task.
- **Landmarks as subgoals:**

  Example: `has-key` > `door-unlock` > `at-right-room` > `at-destination`

- Allow us to leverage incorrect symbolic models.

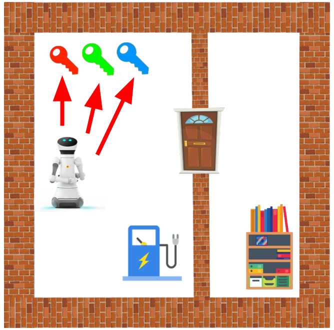
Approximate Symbolic Models Guided RL (ASGRL)

Learning a Diverse Set of Skills Per Subgoal

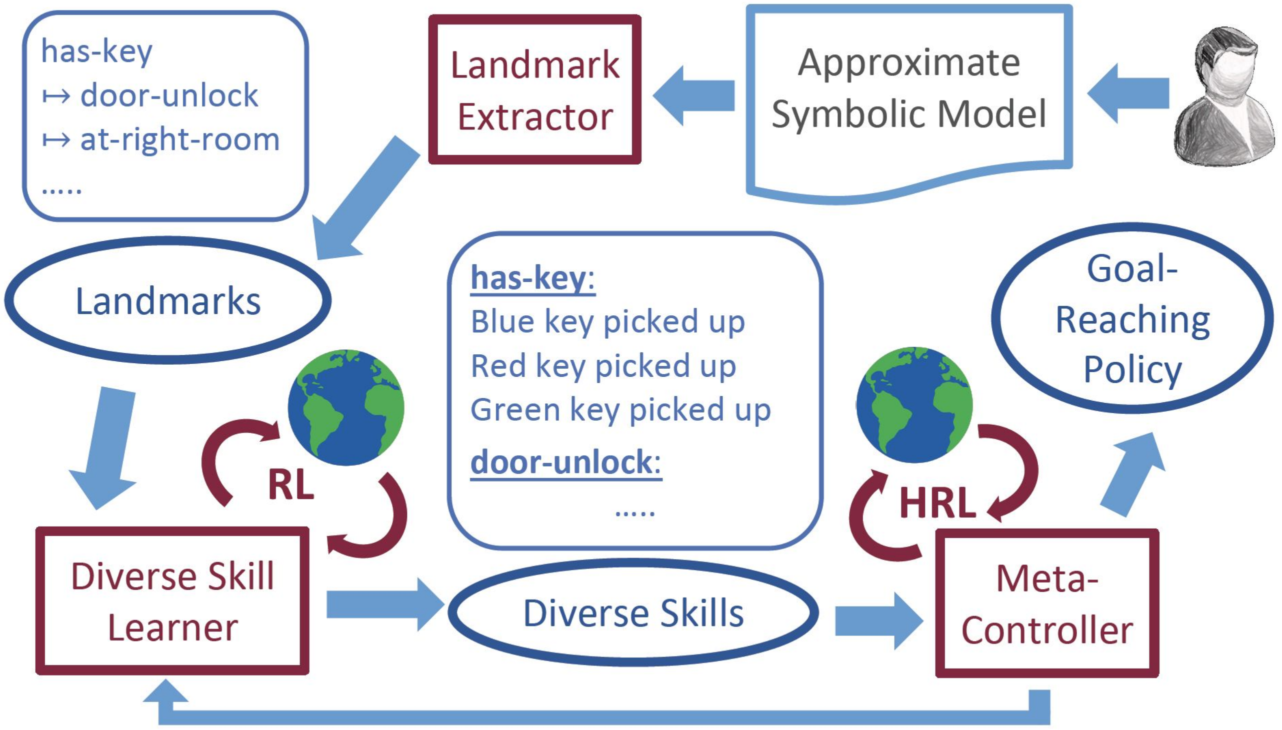
- Given the sources of incompleteness: (a) there could be missing feature(s); (b) one symbolic subgoal state may correspond to a diverse set of low-level states.
- Learn a diverse set of skills to cover different reachable terminal MDP states of a subgoal.
- Can be achieved via an information-theoretic objective: $min\ \mathcal{H}(Z_f|G_f)$

Example:

Subgoal: has-key
Skill 1: pick up red key
Skill 2: pick up green key
Skill 3: pick up blue key

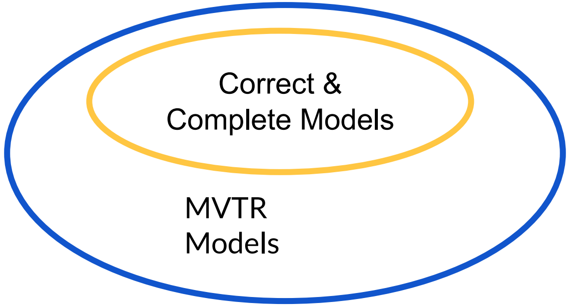


Approximate Symbolic Models Guided RL (ASGRL)



Theoretical Analysis

- ASGRL is guaranteed to result in a goal-reaching policy for all MVTR models
- Minimally Viable Task Representation (MVTR)
  - **MVTR Condition:**
    At least one plan for the symbolic model captures the **relative orderings** of fluents that appear in a low-level goal-reaching trace.
  - A much more **relaxed** condition:
    Individual symbolic action ≠ An executable temporally extended operator at low-level.



Experimental Evaluations

- When inexact and incomplete symbolic models are given, ASGRL manages to efficiently solve the tasks while other baselines fail.
- Three domains and different symbolic models:

	Household-V1	Household-V2	MineCraft	Mario
ASGRL	0.7	**0.9**	**0.9**	**0.9**
ASGRL-Curriculum	**0.8**	**0.9**	**0.9**	**0.9**
Landmark-HRL	0	0	0.1	0
Plan-HRL	0	0	0	0
Landmark-Shaping	0.26	0.43	0.54	0.58
Goal-Q-Learning	0.6	0.6	0.38	0.31