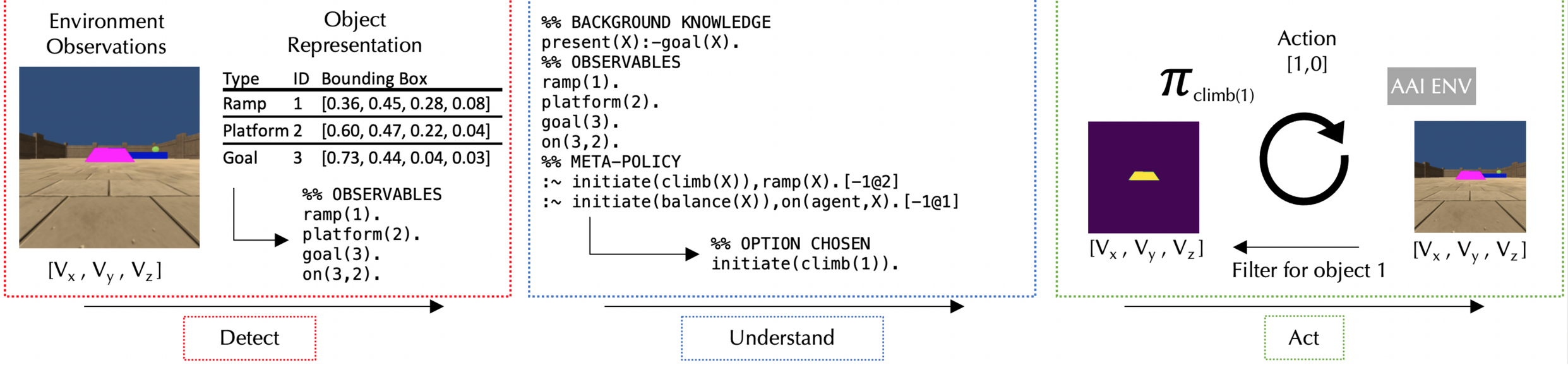
Detect, Understand, Act: A Neuro-Symbolic Hierarchical Reinforcement Learning Framework

Ludovico Mitchener David Tuckey Matthew Crosby Alessandra Russo

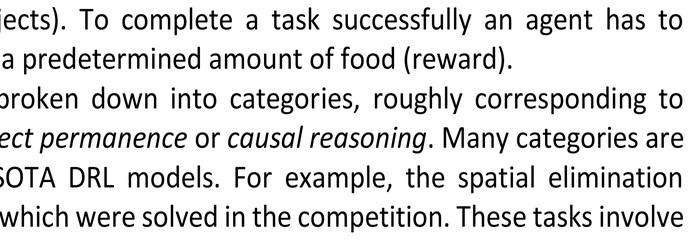
Abstract

In this paper we introduce Detect, Understand, Act (DUA), a neuro-symbolic reinforcement learning framework. The Detect component is composed of a traditional computer vision object detector and tracker. The Act component houses a set of options, high-level actions enacted by pre-trained deep reinforcement learning (DRL) policies. The Understand component provides a novel answer set programming (ASP) paradigm for symbolically implementing a meta-policy over options and effectively learning it using inductive logic programming (ILP). We evaluate our framework on the Animal-AI (AAI) competition testbed, a set of physical cognitive reasoning problems. Given a set of pre-trained DRL policies, DUA requires only a few examples to learn a meta-policy that allows it to improve the state-of-the-art on multiple of the most challenging categories from the testbed. DUA constitutes the first holistic hybrid integration of computer vision, ILP and DRL applied to an AAI-like environment and sets the foundations for further use of ILP in complex DRL challenges.

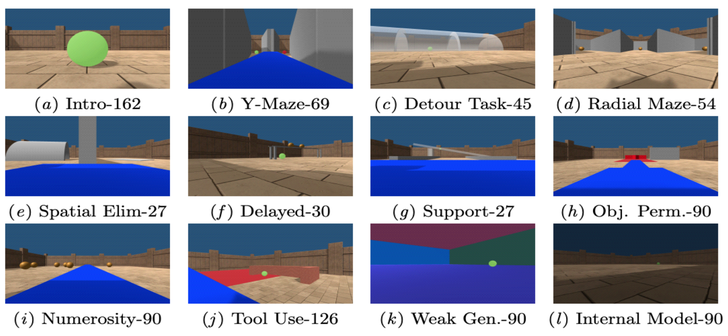


Animal AI

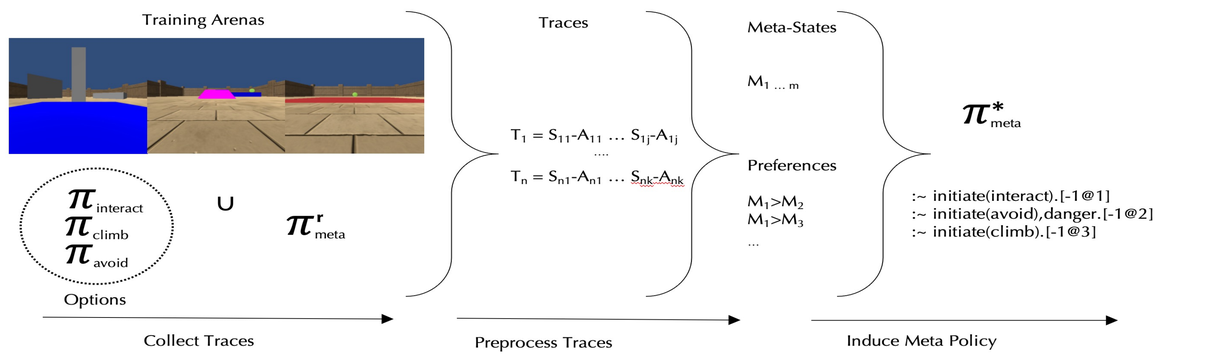
The AAI environment [1] comprises of a small arena in which various objects can be placed to recreate tasks used in animal cognition. To simplify the environment in order to focus on the cognitive abilities being tested, the objects are colour coded and of relatively few base types e.g. walls, ramps and food (reward objects). To complete a task successfully an agent has to navigate the environment to collect a predetermined amount of food (reward).



The testbed consists of 900 tests broken down into categories, roughly corresponding to different cognitive skills, such as *object permanence* or *causal reasoning*. Many categories are incredibly challenging for current SOTA DRL models. For example, the spatial elimination category includes 27 tasks, only 7 of which were solved in the competition. These tasks involve inferring the only possible location that food could be in (behind an opaque object) and directing exploration in that area. These tasks are purposefully designed such that an undirected (e.g. random) exploration strategy will fail. On the other hand, it is not possible to apply symbolic learning methods directly to the environment due to the pixel inputs and low-level control provided by the action space.



Inductive Meta-Policy Learning

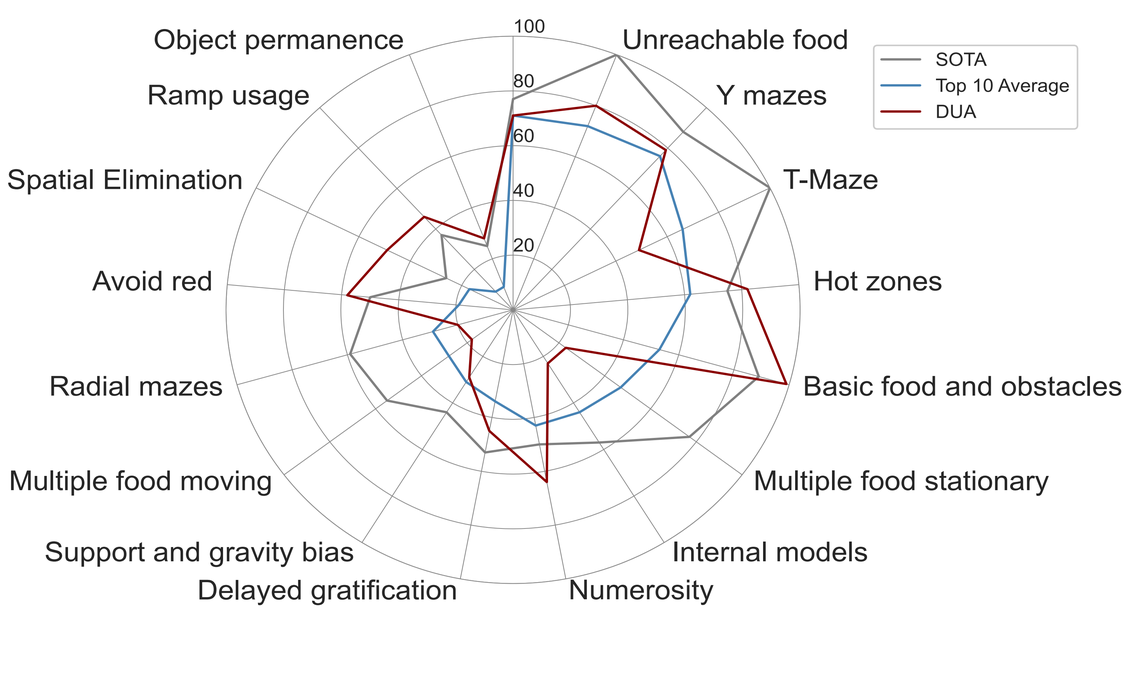


Inductive Meta-Policy learning (IMP) represents the core of our contribution and is our approach for learning a symbolic meta-policy over. In order to learn a meta-policy, we need to create a learning from answer sets task. Meta-policy learning happens in three steps:

1. Collect the meta-traces by running the agent in the environment and at each macro-step randomly picking options to execute π^{r}_{meta} . We store the meta-traces along with their respective episode success in the set of tuples T.
2. We abstract each meta-trace: we map the state-option pairs in the meta-traces in T to a set Ta of tuples including the abstract state-option pairs and associated expected return. This step finds in T similar state-option pairs and combines them to obtain a value akin to a Q-value.
3. We map the generated set Ta into a learning from answer set task Ti to learn the meta-policy π^{*}_{meta}.

Results

To evaluate DUA+IMP on the AAI testbed we implemented 9 options and created 7 training arenas. The final meta-policy learned is displayed below on the right alongside the results obtained on the testbed on the left. We outperform all 60 competitors' submissions on multiple of the most challenging categories. DUA achieves state-of-the-art results in all the categories related to the 7 training arenas, with the exception of y-mazes, where it still outperforms the top 10 average. This suggests that the meta-policy learned is robust and can generalise to a variety of cognitive reasoning tasks outside its training distribution.



Learned Meta-Policy

```
:~ initiate(climb).[-1@11].
:~ danger, initiate(observe),
on(agent,platform).[-1@10].
:~ initiate(drop(V1)), more_goals(V1).[-
1@9, V1].
:~ initiate(collect), not lava.[-1@8].
:~ initiate(interact(V1)), not danger,
not on(goal,platform).[-1@7, V1].
:~ initiate(explore(V1)),
occludes_more(V1,V2).[-1@6, V1, V2].
:~ initiate(explore(V1)),
occludes(V1).[-1@5, V1].
:~ initiate(avoid).[-1@4].
:~ initiate(balance).[-1@3]
:~ bigger(V1,V2),
initiate(interact(V1)).[-1@2, V1, V2].
:~ initiate(rotate).[-1@1].
```

References

[1] Matthew Crosby, Benjamin Beyret, Murray Shanahan, Jose´ Herna´ndez-Orallo, Lucy Cheke, and Marta Halina. The animal-ai testbed and competition. In NeurIPS 2019 Competition and Demonstration Track, pages 164–176. PMLR, 2020.

Imperial College London