Extended Abstract: Non-monotonic Logical Reasoning and Deep Learning for Transparent Decision Making in Robotics

Tiago Mota
Electrical, Computer and Software Engineering, The University of Auckland, NZ
tandrademota@gmail.com

Mohan Sridharan, Ales Leonardis
School of Computer Science, University of Birmingham, UK
m.sridharan@bham.ac.uk, a.leonardis@bham.ac.uk

This abstract summarizes an architecture that enables a robot to provide on-demand *explanations* of its decisions and beliefs [1]. These explanations are in the form of descriptions comprising relations between relevant domain objects, object attributes, robot attributes, and robot actions. Such "explainability" will improve the underlying algorithms, establish accountability, and support more effective human-robot collaboration. However, state of the art robot architectures often include knowledge-based reasoning methods (e.g., for planning) and data-driven learning methods (e.g., for recognizing objects). Providing transparency is challenging in such *integrated robot systems* because the robot has to sense and interact with the physical world, reason with different descriptions of knowledge and uncertainty (e.g., logic-based descriptions of commonsense domain knowledge, probabilistic uncertainty quantification), and incrementally revise its domain knowledge (e.g., axioms governing action effects).

Towards achieving transparency in integrated robot systems, our architecture builds on cognitive systems research that indicates the benefits of formally coupling different representations, reasoning methods, and learning methods. Specifically, it supports the following key capabilities:

- Make decisions reliably and efficiently based on non-monotonic logical reasoning and probabilistic reasoning with incomplete domain knowledge and observations at different resolutions;
- In situations when reasoning is unable to complete the target tasks, automatically identify and use relevant information to learn (deep network) models for these tasks;
- Automatically identify and use relevant information to learn axioms encoding previously unknown domain constraints, and action preconditions and effects;
- Automatically trace the evolution of any given belief or the non-selection of any given action at a given time by inferring the relevant sequence of axioms and beliefs; and
- Exploit the interplay between representation, reasoning, and learning to reliably and efficiently provide on-demand descriptions of decisions and beliefs.

These capabilities are evaluated in simulation and on a robot: (i) computing and executing plans to achieve object configurations; and (ii) estimating occlusion and stability of objects.

Figure 1(left) provides an overview of the architecture. Answer Set Prolog (ASP) is used to encode incomplete commonsense domain knowledge that includes some object attributes, spatial relations between objects; domain attributes; features from images of scenes; and axioms governing domain dynamics. The robot performs non-monotonic logical reasoning with this knowledge to compute plans to achieve any given goal and/or to complete the estimation tasks, using probabilistic reasoning when necessary. If the robot is unsuccessful or achieves an incorrect outcome, this is considered to indicate

Submitted to:
ICLP 2021

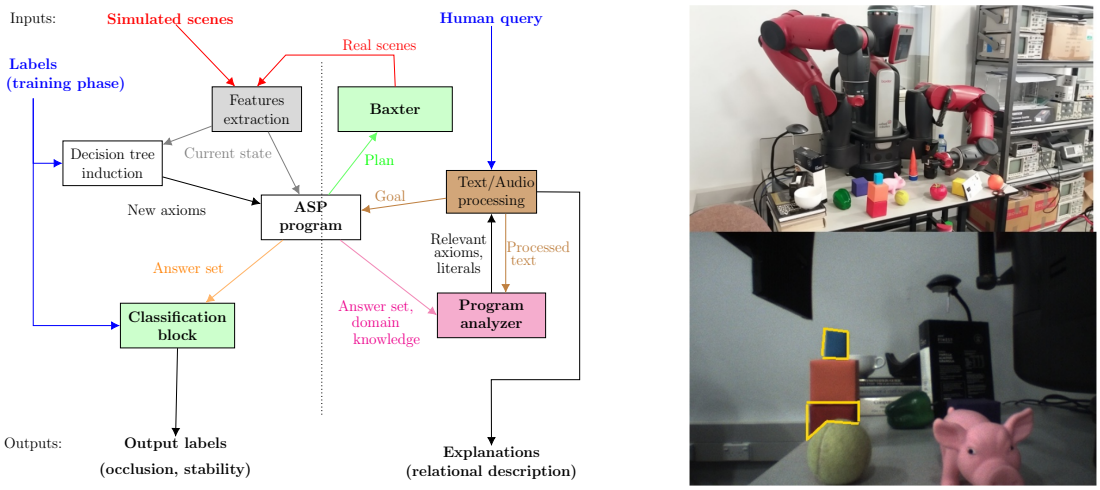

Figure 1: (Left) Overview of the architecture; (Right) Baxter robot setup and example scene.

missing or incorrect knowledge. The robot automatically identifies and uses regions of interest (ROIs) in the relevant images to train and use deep networks to perform these tasks. Information from the ROIs is also used to induce previously unknown axioms and revise existing axioms. The robot also parses the human input to identify the query type (e.g., descriptive, contrastive, counterfactual), and to trace relevant beliefs and axioms to construct relational descriptions that respond to these queries. Experimental results indicate the ability to: (i) make decisions reliably and efficiently despite incomplete knowledge and noisy sensor inputs; (ii) incrementally learn previously unknown constraints, and preconditions and effects of actions; and (iii) accurately explain decisions and beliefs.

As an example, consider the scene in Figure 1 (bottom right). The following interaction takes place *after* the robot has executed a plan to successfully move the red cube on the orange cube.

- **Human:** "Please describe the plan."
  Baxter: "I picked up the blue cube. I put the blue cube on the table. I picked up the orange cube. I put the orange cube on the table. I picked up the red cube. I put the red cube on the orange cube."
- The human may ask the robot to justify a particular action choice.
  Human: "Why did you pick up the blue cube at step 0?"
  Baxter: "Because I had to pick up the red cube, and it was below the blue cube."
  The image regions that influenced this answer are also highlighted—see Figure 1 (bottom right).
- **Human:** "Why did you not put down the orange cube on the blue cube?"
  Baxter: "Because the blue cube is small."
  This answer uses the learned default that a large object on a small object is typically unstable.
- The human may also ask the robot to justify particular beliefs.
  Human: "Why did you believe that the red cube was below the blue cube in the initial state?"
  Baxter: "Because I observed the red cube below the blue cube in step zero."

References

[1] Tiago Mota, Mohan Sridharan & Ales Leonardis (2021): *Integrated Commonsense Reasoning and Deep Learning for Transparent Decision Making in Robotics*. *Springer Nature Computer Science* 2(242), pp. 1–18, doi:10.1007/s42979-021-00573-0.