Cognitive Adequacy: Insights from Developing Robot Architectures⋆

Mohan Sridharan[1],*,†

[1] *Intelligent Robotics Lab, School of Computer Science, University of Birmingham, UK*

Abstract

This paper discusses cognitive adequacy for robots collaborating with and assisting humans. We share insights from the development of robot architectures that use knowledge-driven and data-driven methods to jointly address challenges in transparent knowledge representation, reasoning, and learning in robotics.

Keywords

Non-monotonic logical reasoning, Probabilistic reasoning, Interactive learning, Robotics

1. Motivation

Consider a robot delivering objects to particular places or stacking objects in desired configurations. Such robots have to reason with different descriptions of incomplete domain knowledge and uncertainty. These descriptions include commonsense knowledge, e.g., relations between some domain objects and default statements such as "textbooks are usually in the library" that hold true in all but a few exceptional circumstances. At the same time, information extracted from noisy sensor inputs is often associated with quantitative measures of uncertainty, e.g., "I am 90% certain I saw the robotics book in the office". Also, any robot in a practical domain will have to revise its existing knowledge over time, often using data-driven methods. In addition, for effective collaboration with humans, the robot may need to describe (or justify) its decisions and beliefs. In state of the art architectures that combine knowledge-based reasoning (e.g., for planning) and data-driven learning (e.g., for object recognition) for such *integrated robot systems*, *cognitive adequacy*, i.e., the ability to support the desired behavior, thus poses open problems in knowledge representation, reasoning, and learning. This paper builds on expertise in designing robot architectures [1, 2, 3] to identify some key underlying principles for cognitive adequacy.

2. Architecture and Insights

Figure 1(left) is an overview of the architecture that encodes the principle of *stepwise iterative refinement*. It is based on tightly-coupled transition diagrams at different resolutions, and may be viewed as a logician, statistician, and an explorer working together. These diagrams are described using an action language, which has a sorted signature with statics, (Boolean, non-Boolean) fluents and actions, and supports (deterministic, non-deterministic) causal laws, state constraints,

Workshop on Cognitive Aspects of Knowledge Representation at IJCAI 2022

*Corresponding author.

m.sridharan@bham.ac.uk (M. Sridharan)
https://www.cs.bham.ac.uk/~sridharm/ (M. Sridharan)
0000-0001-9922-8969 (M. Sridharan)

© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

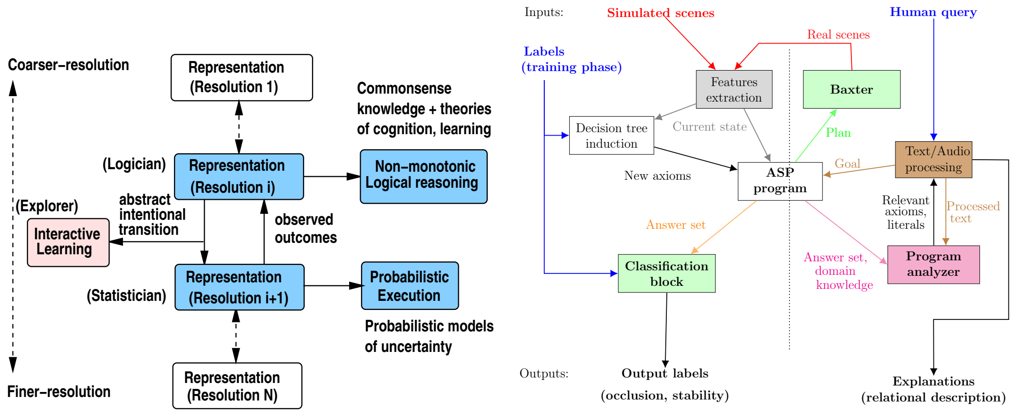
Coarser−resolution
Representation (Resolution 1)
Commonsense knowledge + theories of cognition, learning
(Logician)
Representation (Resolution i)
Non−monotonic Logical reasoning
(Explorer)
abstract intentional transition
observed outcomes
Interactive Learning
(Statistician)
Representation (Resolution i+1)
Probabilistic Execution
Probabilistic models of uncertainty
Finer−resolution
Representation (Resolution N)
Inputs:
Simulated scenes
Human query
Labels (training phase)
Real scenes
Features extraction
Baxter
Decision tree induction
Current state
Plan
Text/Audio processing
New axioms
ASP program
Goal
Relevant axioms, literals
Processed text
Answer set
Answer set, domain knowledge
Classification block
Program analyzer
Outputs:
Output labels (occlusion, stability)
Explanations (relational description)

Figure 1: Architecture represents and reasons with transition diagrams at different resolutions, combining strengths of declarative programming, probabilistic reasoning, and interactive learning.

and executability conditions. The domain's history includes observations, action executions, and prioritized defaults. For any given task, the robot plans and executes actions at two resolutions, but constructs on-demand relational descriptions of decisions and beliefs at other resolutions.

Knowledge representation and reasoning: With two resolutions, the robot represents and reasons with commonsense domain knowledge, including cognitive theories, in the coarse-resolution. For example, a robot fetching objects in an office building reasons about the knowledge it has about some attributes and default room locations of objects. It also has a *adaptive theory of intentions* encoding principles of *non-procrastination* and *persistence*. The fine-resolution transition diagram is defined as a *refinement* of the coarse-resolution transition diagram, introducing a *theory of observations* that models the robot's ability to sense the values of domain fluents. A robot in an office building would (for example) now consider grid cells in rooms and object parts, attributes that were previously abstracted away by the designer. The definition of refinement guarantees that for any given coarse-resolution transition, there exists a path in the fine-resolution diagram between states that are refinements of the coarse-resolution states. Also, the refined diagram is *randomized* to model non-determinism in action outcomes. For any given goal, non-monotonic logical reasoning at the coarse-resolution provides a plan of *intentional abstract actions*; this is achieved using *Answer Set Prolog*, a declarative programming paradigm [4]. The robot implements each abstract transition as a sequence of concrete actions by automatically identifying (i.e., *zooming* to) and reasoning with the *relevant* part of the fine-resolution diagram. Execution in the fine-resolution uses probabilistic models of the uncertainty (e.g., in perception, actuation), with the outcomes added to the coarse-resolution history for subsequent reasoning [1, 2].

Interactive learning and transparency: Reasoning with incomplete knowledge can result in incorrect or suboptimal outcomes. State of the art machine learning methods (e.g., using deep networks) require many labeled examples and considerable computational resources that are often not available in practical robot domains. The architecture supports three strategies for incremental, efficient acquisition of knowledge of previously unknown action capabilities and axioms: (i) verbal descriptions of observed behavior; (ii) exploration of new transitions; and (iii) reactive exploration of unexpected transitions. These strategies are formulated as suitable interactive (e.g., inductive, reinforcement) learning problems. *Reasoning and learning guide each* *other*, enabling the automatic identification and use of only the relevant information to construct mathematical models for the different formulations [5]. For example, to estimate the stability of objects in a scene, the robot first attempts to reason with domain knowledge and information (e.g., object category, spatial relations) extracted from input images. Relevant regions of interest are automatically extracted from images for which reasoning is unable to make a decision (or makes an incorrect decision), and used to train a deep network. Information from these regions is also used to induce axioms used for subsequent reasoning—Figure 1(right). This approach substantially improves reliability and efficiency in comparison with deep network methods [3, 6].

The architecture supports transparent reasoning and learning, i.e., *explainable agency*, by encoding a *theory of explanations* comprising: (i) claims about representing, reasoning with, and learning knowledge to support relational descriptions of decisions and beliefs; (ii) a characterization of explanations based on representational abstraction, and explanation specificity and verbosity; and (iii) a methodology for constructing explanations. This theory is implemented in conjunction with the components summarized above—see Figure 1(right). The robot then provides on-demand relational descriptions of decisions and beliefs in response to different types of questions (e.g., descriptive, contrastive, counterfactual) posed by a human. The human is able to interactively obtain descriptions at the desired abstraction, specificity, and verbosity, with the robot posing disambiguation questions to the human as needed [3, 7].

3. Execution Trace

The following execution traces demonstrate the working of the architecture.

Execution Example 1. *[Planning and Learning]*
Consider a robot in a $study$ that is asked to fetch a cup.

- The plan of abstract actions: $move(rob_1, kitchen)$, $pickup(rob_1, C)$, $move(rob_1, study)$, $putdown(rob_1, C)$, is based on the default knowledge that cups are usually in the $kitchen$.
- For each abstract transition, the relevant (zoomed) fine-resolution description is identified, e.g., only cells in the $study$ and the $kitchen$ are relevant to the first $move$, and used to obtain a probabilistic policy that is invoked repeatedly to execute a sequence of concrete actions, e.g., robot is in a cell in the $kitchen$ after first $move$.
- Suppose the robot's attempt to pick up a cup in the kitchen fails. Using the knowledge that the cup is $heavy$ and its its arm is $light$, the robot learns the axiom: $\mathbf{impossible}$ $pickup(rob_1, C)$ **if** $arm(rob_1, light)$, $obj\_weight(C, heavy)$,
- When asked to provide a detailed description after plan execution, the robot revises the abstraction level to use the fine-resolution description.
  Human: "Please describe the executed plan in detail."
  Robot: "I moved to cell c_2 in the $kitchen$. I picked the large cup by its handle from the counter [...] I moved to cell c_4 of the $study$. I put the cup down on the red table."

Execution Example 2. *[Explanation and Disambiguation]*
In the simulated scenario in Figure 2, the human asks the robot to "Move the yellow object on the green cube.". The reference to yellow object is ambiguous, and the robot asks for clarification.

- **Robot**: "Should I move yellow duck on the green cube?"
  Human: "No. Move yellow cylinder on the green cube."

- The robot computes a plan: *pick up green mug; put green mug on table; pick up red cube; put red cube on table; pick up yellow cube; put yellow cube on table; pick up yellow cylinder; put yellow cylinder on green cube.*
- The robot traces beliefs and axioms to answer questions after plan execution.
  Human: "Why did you not pick up red cube at step1?"
  Robot: "Because the red cube was below the green mug."
  Human: "Why did you move yellow cube to the table?"
  Robot: "I had to put the yellow cylinder on top of the green cube. The green cube was below the yellow cube."

Summary: Implementing principles such as stepwise iterative refinement and relevance, and exploiting the interplay between representation, reasoning, and learning, are key steps towards achieving cognitive adequacy in architectures for robots. Such an architecture that combines knowledge-based reasoning and data-driven learning has providing promising results in simulation and on physical robots [1, 2, 3, 5].

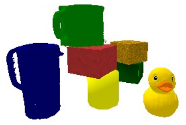

Figure 2: Example.

Acknowledgments

This work is the result of research threads pursued in collaboration with Ben Meadows, Tiago Mota, Heather Riley, Rocio Gomez, Michael Gelfond, Shiqi Zhang, and Jeremy Wyatt. This work was supported in part by the U.S. ONR Awards N00014-13-1-0766, N00014-17-1-2434 and N00014-20-1-2390, AOARD award FA2386-16-1-4071, and U.K. EPSRC award EP/S032487/1.

References

[1] R. Gomez, M. Sridharan, H. Riley, What do you really want to do? Towards a Theory of Intentions for Human-Robot Collaboration, Annals of Mathematics and Artificial Intelligence, special issue on commonsense reasoning 89 (2021) 179–208.

[2] M. Sridharan, M. Gelfond, S. Zhang, J. Wyatt, REBA: A Refinement-Based Architecture for Knowledge Representation and Reasoning in Robotics, Journal of Artificial Intelligence Research 65 (2019) 87–180.

[3] T. Mota, M. Sridharan, A. Leonardis, Integrated Commonsense Reasoning and Deep Learning for Transparent Decision Making in Robotics, Springer Nature CS 2 (2021) 1–18.

[4] M. Gebser, R. Kaminski, B. Kaufmann, T. Schaub, Answer Set Solving in Practice, Synthesis Lectures on Artificial Intelligence and Machine Learning, Morgan Claypool Publishers, 2012.

[5] M. Sridharan, B. Meadows, Knowledge Representation and Interactive Learning of Domain Knowledge for Human-Robot Collaboration, Advances in Cognitive Systems 7 (2018) 77–96.

[6] H. Riley, M. Sridharan, Integrating Non-monotonic Logical Reasoning and Inductive Learning With Deep Learning for Explainable Visual Question Answering, Frontiers in Robotics and AI, special issue on Combining Symbolic Reasoning and Data-Driven Learning for Decision-Making 6 (2019) 20.

[7] M. Sridharan, B. Meadows, Towards a Theory of Explanations for Human-Robot Collaboration, Kunstliche Intelligenz 33 (2019) 331–342.