Knowledge Representation, Explainable Reasoning and Interactive Learning in Robotics

ICAPS Journal Paper Presentation Track, 2020

Mohan Sridharan[1]

Intelligent Robotics Lab
School of Computer Science, University of Birmingham, UK
m.sridharan@bham.ac.uk
https://www.cs.bham.ac.uk/~sridharm/

October 27/30, 2020

[1] Ben Meadows, Rocio Gomez, Heather Riley, Tiago Mota (Univ. of Auckland, NZ); Michael Gelfond (Texas Tech, USA); Jeremy Wyatt (Univ. of Birmingham, UK); Shiqi Zhang (SUNY Binghamton, USA); Pat Langley (ISLE, USA), US ONR Awards N00014-13-1-0766, N00014-17-1-2434; US AFOSR/AOARD Award FA2386-16-1-4071; EPSRC/EU projects

Research Questions

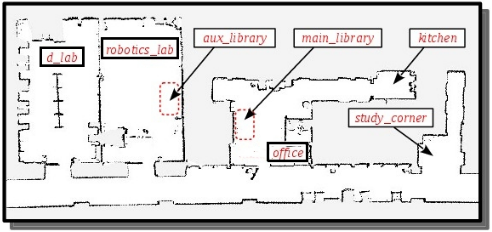
- How best to enable robots to represent and reason with qualitative and quantitative descriptions of incomplete knowledge and uncertainty?
  "Books are usually in the library"
  "I am 90% certain the robotics book is in the library"
- How best to enable robots to learn interactively and cumulatively from sensor inputs and limited human feedback.
  Learn actions, action capabilities, domain dynamics
  "Robot with weak arm cannot lift heavy box"
- How to enable designers to understand the robot's behavior and establish that it satisfies desirable properties.
  Explainable agency, intentions, goals, measures
  "What would happen if I dropped the the spoon on the table?"

Inspiration and Core Ideas

- Cognitive systems inspired by human cognition and control.
- Represent, reason, learn jointly at different abstractions with different schemes (Alan Turing, 1952; morphogenesis).
- Logician, statistician, and creative explorer; tight coupling not unified representation (Immanuel Kant, Aaron Sloman).
- Principle of stepwise iterative refinement (Edsger Dijkstra).
- Interactive and cumulative learning of relevant concepts.
- *Not focusing on hardware, energy requirements*.

Illustrative Domain: Robot Assistant

Robot assistant finding and manipulating objects.



Claims: Representation

1. Distributed representation of knowledge (commonsense, probabilistic) at different abstractions.
2. Knowledge structures include definitions, constraints.
3. Beliefs include prior knowledge, inferences, plans, explanations.
4. History includes observations, actions (+ defaults?).
5. Separation of concerns (domain-specific/independent knowledge, observations), but abstractions tightly coupled.
6. Possible worlds, each a set of beliefs.

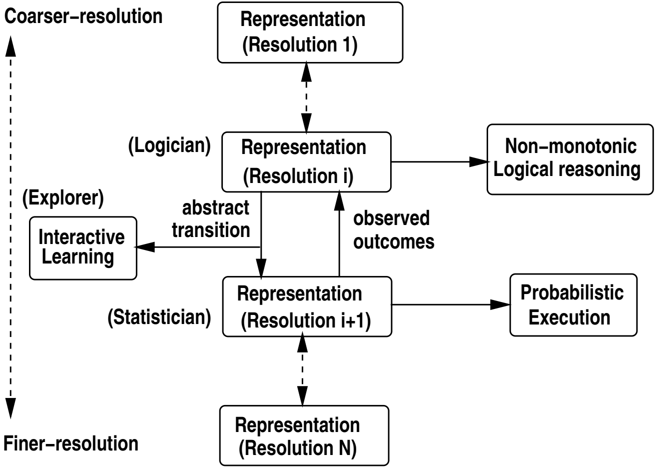
Claims: Reasoning

1. Knowledge elements support non-monotonic revision; revise previously held conclusions.
2. Actions produce immediate or delayed outcomes; reward-based and architecture-based exploration.
3. Observations obtained through active exploration or reactive action execution.
4. "Here and there" reasoning; satisfiability, stochastic policies.

Action Language AL_d

- Formal models of parts of natural language used for describing transition diagrams.
- Hierarchy of basic sorts, statics, fluents and actions.
- Types of statements:
  - Causal law (deterministic, non-deterministic).
  - State constraint and definitions.
  - Executability condition.

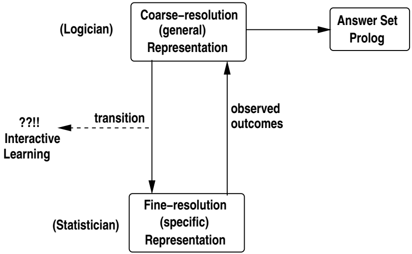
Refinement-Based Architecture (REBA)



Mohan Sridharan, Michael Gelfond, Shiqi Zhang and Jeremy Wyatt. **REBA: Refinement-based Architecture for Knowledge Representation and Reasoning in Robotics**. In Journal of Artificial Intelligence Research, 65:87-180, May 2019.

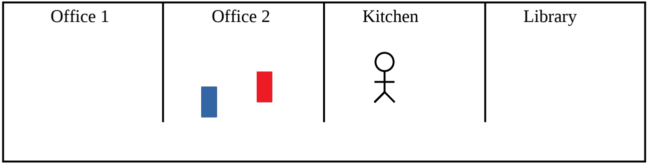
Logician's Description and Reasoning

- Logician's description:
  - **Inputs**: (a) $\mathcal{D}_H$ as AL_d statements of sorted signature and axioms; (b) history $\mathcal{H}$ with initial state defaults; (c) Goal.
  - **Output**: plan of transitions to execute.



- Construct Answer Set Prolog program $\Pi(\mathcal{D}_H, \mathcal{H})$. Reasoning reduced to computing answer sets.

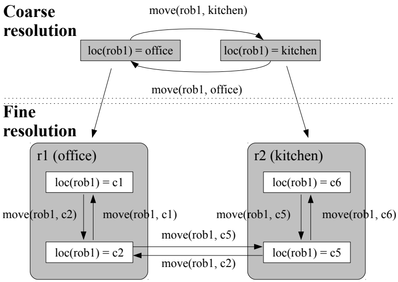
Theory of Intentions: Motivation

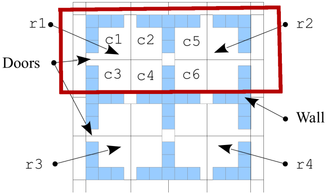


- Unexpected success and failure.
- **Approach:** model intention and related observations.
  - Persistence and non-procrastination (Blount and Gelfond, 2015).
  - Activity, mental fluents and actions.
  - Scaling using relevance and abstraction: $\Pi(\mathcal{D}'_H, \mathcal{H}')$.

Rocio Gomez, Mohan Sridharan, and Heather Riley. **What do you really want to do? Towards a Theory of Intentions for Human-Robot Collaboration**. In Annals of Mathematics and Artificial Intelligence, special issue on Commonsense Reasoning, 2020.

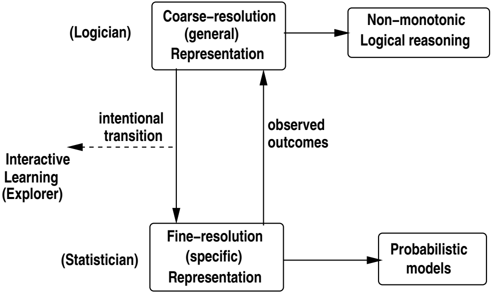
Refine + Zoom + Randomize





- Refinement: describe ($\mathcal{D}_H$) at finer resolution ($\mathcal{D}_L$).
- Theory of observation: knowledge fluents + actions.
- Randomize and zoom to $\mathcal{D}_{LR}(T)$ for $T = \langle \sigma_1, a^H, \sigma_2 \rangle$.
- Formal relationships between descriptions.

Construct and Solve Probabilistic Models

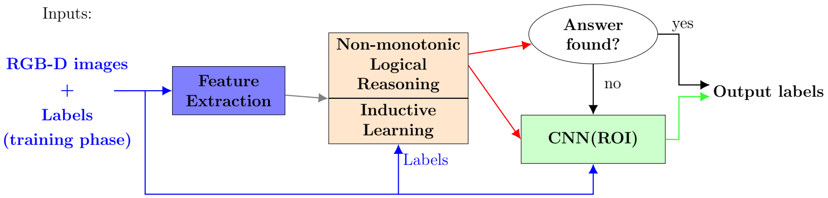


- $\mathcal{D}_{LR}(T)$ and statistics to construct hierarchical probabilistic graphical models, e.g., partially observable Markov Decision Process (POMDP) tuple $\langle S^L, A^L, Z^L, T^L, O^L, R^L \rangle$.
- Add observed outcomes to $\mathcal{H}$ to be used by logician.

Reasoning + Learning + Explanation

- Deep networks widely used in AI and robotics.
  - Large labeled datasets; considerable computational resources; and
  - Representations and mechanisms difficult to interpret.
- Inspiration from human cognition and cognitive systems:
  - Representation, reasoning, learning inform and guide each other.
  - Scalability: abstraction, relevance, and persistence.
  - **Focus:** exploit strengths of non-monotonic logical reasoning, deep learning, and tree induction.
- Relational descriptions of decisions, beliefs, and experiences; in terms of abstraction, specificity, verbosity.
- Experimental domains:
  - Estimate object occlusion, stability; minimize clutter.
  - Answer explanatory questions (VQA) with limited data.

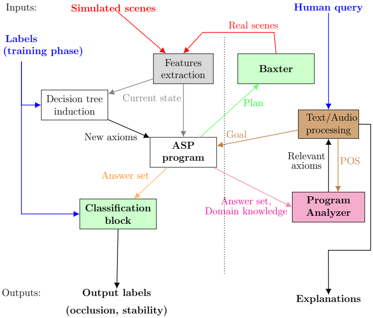
Architecture Components: Overview



Tiago Mota and Mohan Sridharan. **Incrementally Grounding Expressions for Spatial Relations between Objects**. In the International Joint Conference on Artificial Intelligence (IJCAI), July 13-19, 2018.

Tiago Mota and Mohan Sridharan. **Commonsense Reasoning and Knowledge Acquisition to Guide Deep Learning on Robots**. In the Robotics Science and Systems Conference (RSS), Freiburg, Germany, June 22-26, 2019
(Best Paper Award Finalist)

Architecture with Explanations



- Question/request types: (i) describe plan; (ii) why action X? (iii) why not action Y? and (iv) why belief Z?

Mohan Sridharan and Ben Meadows. **Towards a Theory of Explanations for Human-Robot Collaboration**. In Künstliche Intelligenz Journal, 33(4):331-342, December 2019.

Tiago Mota and Mohan Sridharan. **Axiom Learning and Belief Tracing for Transparent Decision Making in Robotics**. In AAAI Fall Symposium on Trust and Explainability in Artificial Intelligence for Human-Robot Interaction, November 2020.

Illustrative Example

- **Goal:** some cup C has to be in the office:
  $loc(C) = office, \;\; \neg in\_hand(rob_1, C)$.

- **Initial knowledge** (subset): $loc(rob_1, office)$,
  $obj\_weight(cup_1, heavy), arm\_type(rob_1, electromagnetic)$.

- Based on **default**: $loc(cup_1) = kitchen$.

- One possible plan from ASP-based inference:

$$move(rob_1, kitchen), \;\; grasp(rob_1, cup_1)$$
$$move(rob_1, office), \;\; putdown(rob_1, cup_1)$$

- Assume rob_1 is in $kitchen$. Has to locate and grasp cup_1.

Illustrative Example (contd.)

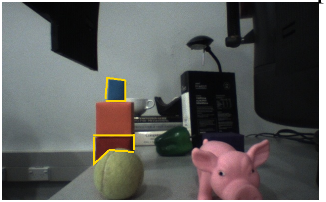
- Some **relevant** literals: $loc(rob_1) = c_i$, $loc(cup_1) = c_j$, where $c_i, c_j \in kitchen$.
- Possible action sequence (executed probabilistically):

$$
\begin{array}{ll}
move(rob_1, c_3) & \\
test(rob_1, loc(cup_1), c_3) & \%\ cup_1 \text{ not observed} \\
move(rob_1, c_5) & \\
test(rob_1, loc(cup_1), c_5) & \%\ cup_1 \text{ observed} \\
grasp(rob_1, cup_1) &
\end{array}
$$

- **Interactive learning** when necessary.

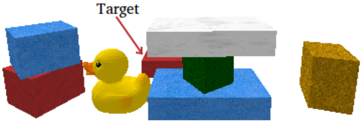
Execution Trace of Explanations

- **Goal**: red block on the top of orange block.



- **Human**: “Why did you pick up the blue block first?”;
- **Baxter**: “Because I had to pick up the red block, and it was below the blue block”;
- **Human**: “Why did you not pick up the orange block first?”;
- **Baxter**: “Because the blue block was on the orange block”;
- **Human**: “What would happen if the ball is pushed?”
- **Robot:** …

Experimental Results: VQA + Decision making



- Accuracy increases and training complexity decreases.
- High precision and recall for learning previously unknown axioms.
- High precision and recall for retrieving relevant literals and constructing explanations.

Heather Riley and Mohan Sridharan. **Integrating Non-monotonic Logical Reasoning and Inductive Learning With Deep Learning for Explainable Visual Question Answering**. In Frontiers in Robotics and AI, special issue on Combining Symbolic Reasoning and Data-Driven Learning for Decision-Making, Volume 6, December 2019.

Contributions

- Step-wise refinement simplifies design and implementation, increases confidence in behavior, promotes scalability.
- Separation of concerns: domain-independent/specific knowledge.
- Non-monotonic logical reasoning, inductive learning, and deep learning inform and guide each other.
- Learned axioms improve decision-making accuracy; explain behavior of deep learning models.
- Interactive explanations constructed efficiently and on demand.