Refinement-based Architecture for Knowledge Representation, Explainable Reasoning, and Interactive Learning in Robotics

Mohan Sridharan, University of Birmingham, UK

m.sridharan@bham.ac.uk

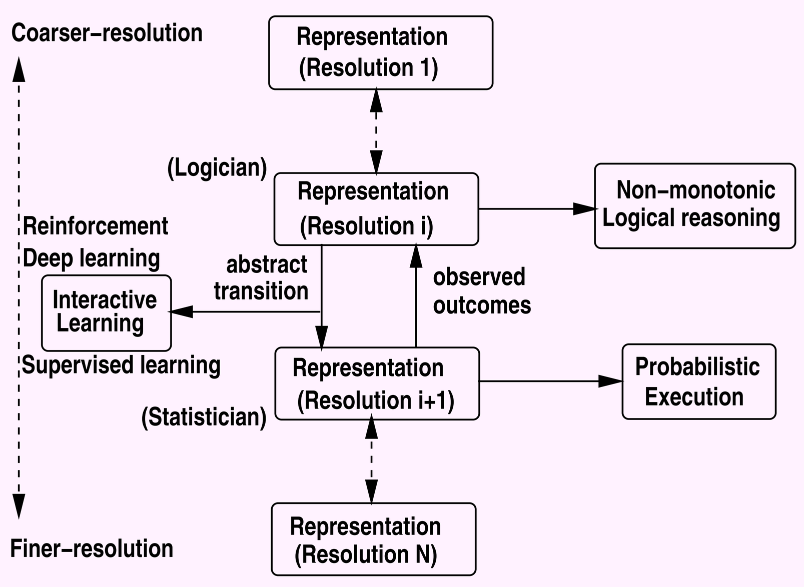
Research Questions

- How best to enable robots to represent and reason with qualitative and quantitative descriptions of incomplete domain knowledge and uncertainty?
  "Books are usually in the library"
  "I am 90% certain robotics book is in the library"
- How best to enable robots to learn interactively and cumulatively from sensor inputs and limited human feedback?
  Camera images, verbal cues, different surfaces
  "Robot with weak arm cannot lift heavy box"
- How best to enable designers to understand the robots' behavior and to establish that it satisfies desirable properties?
  "Why did you go to the kitchen?"
  "How likely is it that the engineer is in the office?

Core Ideas and Inspiration

- Theories of human/animal cognition and motor control.
- Theories of intention, affordance, explanation, observation.
- Qualitative and quantitative reasoning with incomplete knowledge at different abstractions; tight coupling between logician and statistician.
- Interactive and cumulative learning of relevant concepts.

Architecture Overview

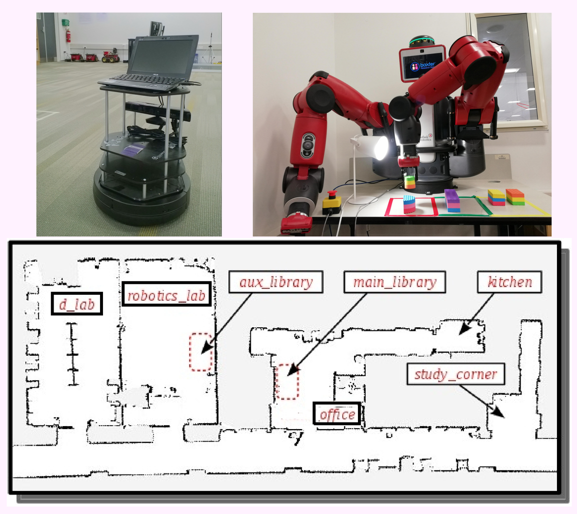


Architecture combines strengths of non-monotonic logical reasoning, probabilistic reasoning, and interactive learning.

Illustrative Domain

Robot Assistant (RA) domain:

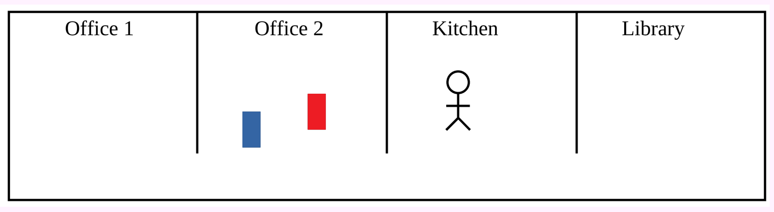
- Find and move target objects (book, cup, printer) to particular places (library, office, workshop, kitchen) or people.
- Humans have *role* (engineer, manager); object attributes color, shape, size.
- Estimate occlusion of objects, stability of structures.
- Answer questions in simulated and real-world scenarios.



Coarse Resolution Domain Representation

- Transition diagrams described as statements of $\mathcal{AL}_d$.
- System description $\mathcal{D}$ with sorted signature Σ and axioms.
- Σ has sorts, statics, and fluents. For RA domain:
  $next\_to(place, place),\ loc(thing) = place,\ stable(object),$
  $in\_hand(robot, object), obj\_relation(relation, object, object)$
- Σ has actions. For RA domain:
  $move(robot, place),\ pickup(robot, object),$
  $putdown(robot, object), exo\_move(object, place)$
- Axioms: constraints, causal laws, executability conditions,
  $move(rob_1, Pl)$ **causes** $loc(rob_1) = Pl$
  $loc(O) = Pl$ **if** $loc(rob_1) = Pl,\ \ in\_hand(rob_1, O)$
  $obj\_relation(above, A, B), I)$ **if** $obj\_relation(below, B, A), I)$
  impossible $pickup(rob_1, O)$ **if** $loc(rob_1) \neq loc(O)$
- History $\mathcal{H}$ with prioritized defaults in initial state.
  initial default $loc(X) = library$ **if** $book(X)$
  initial default $loc(X) = office$ **if** $book(X),$
  $loc(X) \neq library$
- Compute answer sets of CR-Prolog program $\Pi(\mathcal{D}, \mathcal{H})$.
- Non-monotonic logical reasoning essential for robotics+AI.

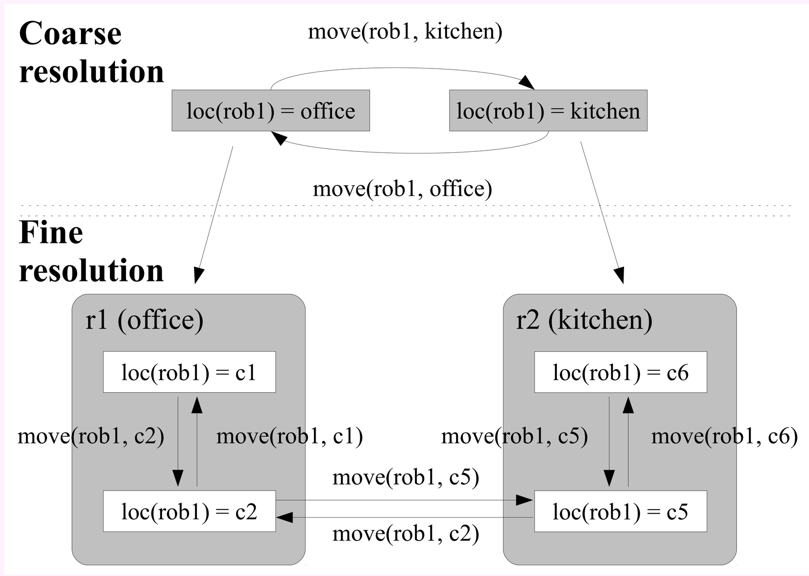
Theory of Affordances and Intentions

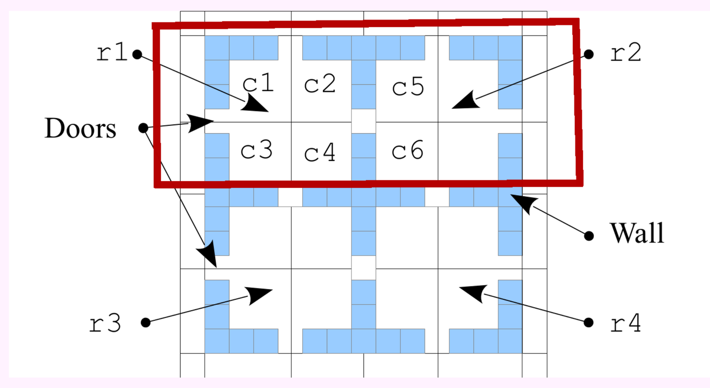


- Affordance: attributes of objects, agents, in terms of actions.
- Unexpected success/failure; intentional actions, observations.
- Principles of persistence, non-procrastination, and relevance.
- Expand $\mathcal{D}$ and $\mathcal{H}$; mental fluents and actions; axioms for action effects, start/stop activities; model attempted actions.

Fine Resolution Domain Representation

- 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.





- Separation of domain-independent/specific knowledge.
- Probabilistic model of uncertainty in sensing and actuation.
- Fine-resolution execution with $\mathcal{D}_{LR}(T)$ and probabilities, e.g., POMDP policy. Add coarse-resolution outcomes to $\mathcal{H}$.

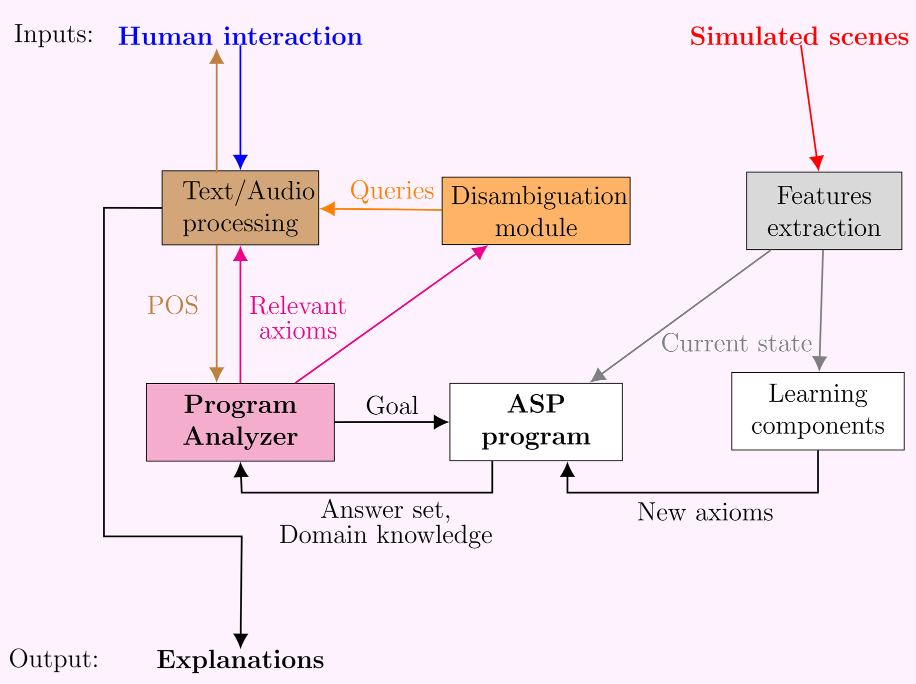
Interactive Learning

- Incomplete knowledge; unexpected or sub-optimal outcomes.
- Labeled samples; limited human time and expertise; actions with delayed outcomes.
- Incrementally learn previously unknown actions, axioms.
- Verbal descriptions from humans: "Robot is labeling fairly big textbook", "Robot labeled small fragile cup"

$$label(R, O)\ \textbf{causes}\ labeled(O)$$

- Relevance and relational inference guide active exploration or reactive execution with knowledge or reinforcement.
- Represent experiences relationally (binary decision tree); cumulative learning and construct new axioms.

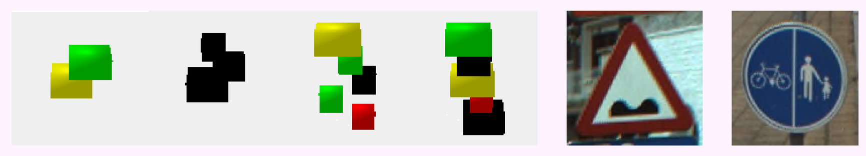
$$\neg stable(A)\ \textbf{if}\ relation(above, A, B),\ surface(B, irregular)$$
$$\textbf{impossible}\ grasp(rob_1, C)\ \textbf{if}\ weight(C, heavy),\ arm(rob_1, electro)$$



Architecture combines non-monotonic logical reasoning, deep learning, and interactive learning, with reasoning and learning informing and guiding each other.

Theory of Explanations and VQA

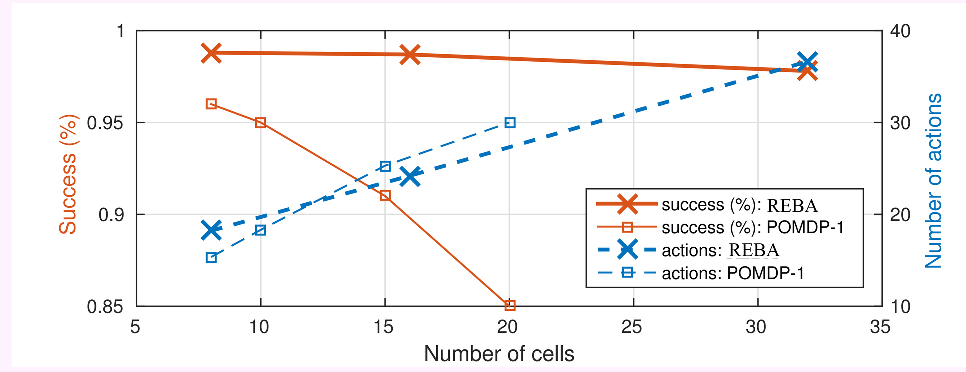
- Characterize explanations: abstraction, specificity, verbosity.
- Methodology for constructing explanations interactively.
- Visual Question Answering (VQA).



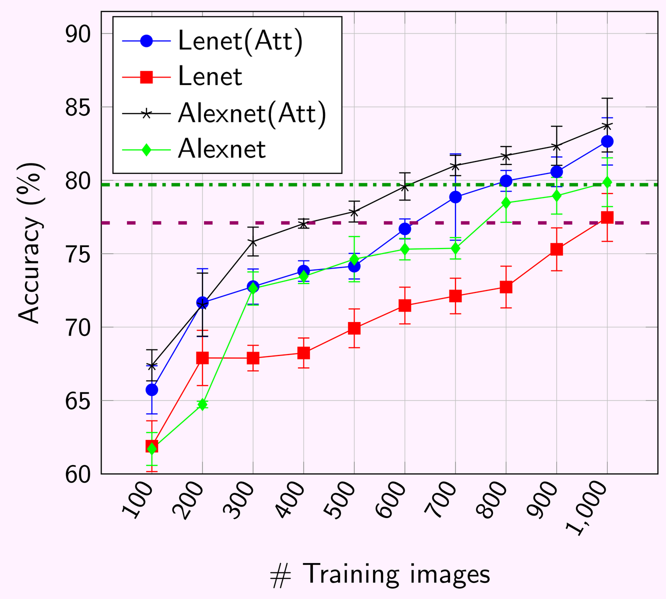
- Complementary strengths of non-monotonic logical reasoning, deep learning, and inductive learning.

Experimental Results

1. Increases accuracy and reduces number of actions executed.



2. Desired accuracy (stability, occlusion) with smaller dataset.



3. Actions and axioms learned with high precision and recall.

Action	Recall	Precision	Axioms	Recall	Precision
label	0.92	0.96	Normal	100%	98%
serve	0.88	0.95	Default	62%	78%

4. Minimal and correct plans with learned knowledge.

Conclusions + Future Work

- Step-wise refinement simplifies design and implementation, increases confidence in behavior, promotes scalability.
- Precise relationship between descriptions at different resolutions.
- Reasoning directs interactive learning of domain dynamics.
- Explanations at desired level of abstraction.
- Further explore interplay between reasoning and learning.

Acknowledgements

REBA-KRL architecture is the result of research threads pursued in collaboration with Ben Meadows, Rocio Gomez, Tiago Mota, Heather Riley, Michael Gelfond, Shiqi Zhang, Jeremy Wyatt, and Pat Langley. Supported by U.S. Office of Naval Research Science of Autonomy Awards N00014-13-1-0766 and N00014-17-1-2434, and Asian Office of Aerospace Research and Development award FA2386-16-1-4071. Conclusions are those of the author.

References

1. 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, 2019.
2. Tiago Mota and Mohan Sridharan. **Commonsense Reasoning and Knowledge Acquisition to Guide Deep Learning on Robots**. In the *Robotics Science and Systems Conference*, Freiburg, Germany, 2019.
3. Mohan Sridharan and Ben Meadows. **Knowledge Representation and Interactive Learning of Domain Knowledge for Human-Robot Interaction**. In *Advances in Cognitive Systems Journal*, 7:77-96, 2018.
4. 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*, 2020.
5. Mohan Sridharan and Ben Meadows. **Theory of Explanation for Human-Robot Collaboration**. In the *Knstliche Intelligenz Journal*, 33(4):331-342, December 2019.