Commonsense Reasoning and Deep Learning for Transparent Decision Making in Robotics[1]

XLoKR at KR-2020

Tiago Mota [a], Mohan Sridharan [b], and Ales Leonardis [b]

[a] Department of Electrical and Computer Engineering
The University of Auckland, NZ

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

September 13, 2020

[1] ONR Awards N00014-13-1-0766, N00014-17-1-2434; AFOSR Award FA2386-16-1-4071

Research Questions

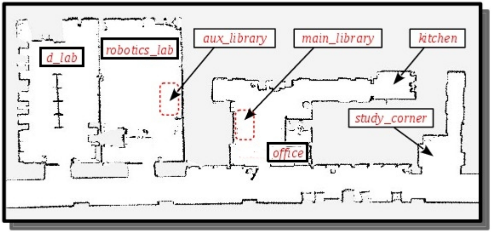
- How best to enable robots to represent and reason with qualitative and quantitative descriptions of incomplete knowledge and uncertainty?
  "Books are usually in the library"
  "I am 90% certain the robotics book is in the library"

- 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?"

Illustrative Domain: Robot Assistant

Robot assistant finding and manipulating objects.



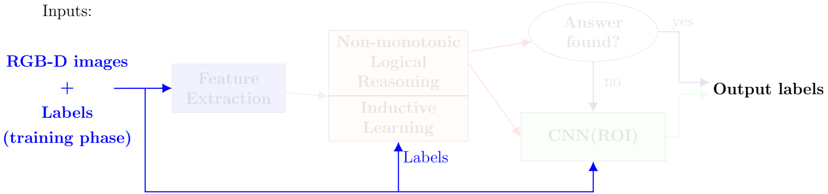
Reasoning + Learning: Motivation

- Deep networks widely used in AI and robotics.
- Limitations of deep network architectures:
  - 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.
- Experimental domains:
  - Estimate object occlusion, stability; minimize clutter.
  - Answer explanatory questions (VQA) with limited data.

Theory + Approach

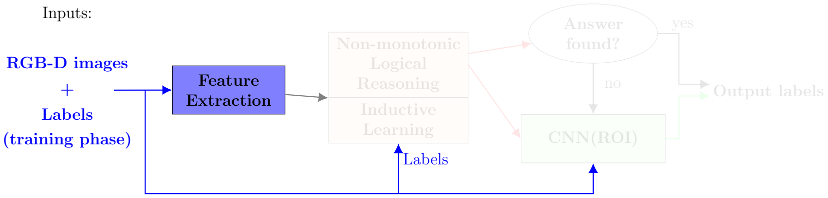
1. Relational descriptions of decisions (+rationale), beliefs, and experiences in terms of domain+agent attributes, actions.
2. Characterize explanations in terms of abstraction, specificity, verbosity; provide methodology to construct explanations.
3. Present contextual/relevant information on-demand and during/after task performance.
4. **Focus:** KR tools for transparency in reasoning and learning; not discussing human studies.
5. Exploit complementary strengths of non-monotonic logical reasoning, deep learning, and decision tree induction.

Architecture Components: Input



- Images: images of objects, scenes.
- Labels: object occlusion, stability of structures, answers.

Architecture Components: Feature Extraction

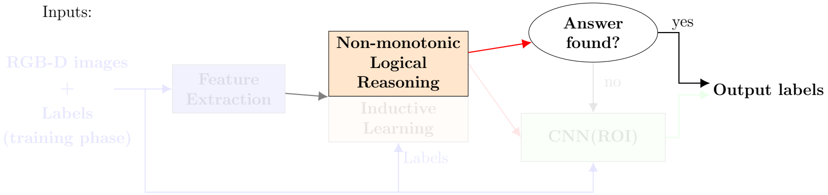


Geometric features extracted from images:

- Spatial relations between objects (above, behind, right of ...).
- Shape and size of objects in the scene.

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.

Architecture Components: Non-monotonic Logic

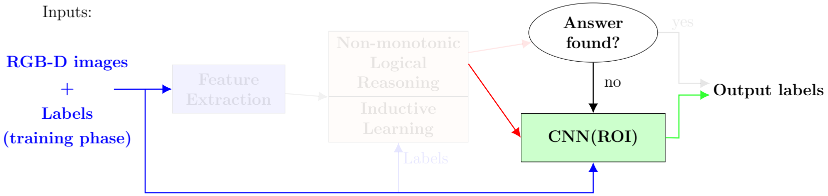


- Input: Extracted features, incomplete domain dynamics.
- ASP for non-monotonic logical reasoning.

$$holds(obj\_rel(on, Ob_1, Ob_2), I+1) \leftarrow occurs(putdown(rob_1, Ob_1, Ob_2), I)$$
$$holds(obj\_rel(above, Ob_1, Ob_2), I) \leftarrow holds(obj\_rel(below, Ob_2, Ob_1), I)$$
$$\neg occurs(pickup(rob_1, Ob_1), I) \leftarrow holds(obj\_rel(below, Ob_1, Ob_2), I)$$

- Decision about input image if possible.

Architecture Components: CNN



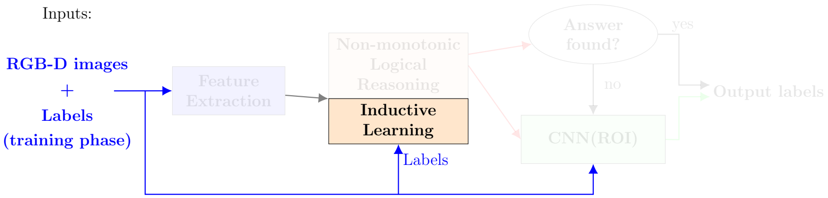
- Attention: ROI selection based on state constraints.

$$stable(A) \leftarrow \neg obj\_rel(above, A, B)$$

$$\neg stable(A) \leftarrow obj\_rel(above, A, B),\ obj\_surface(B, irregular)$$

- CNN: Convolutional Neural Network (Lenet and Alexnet).

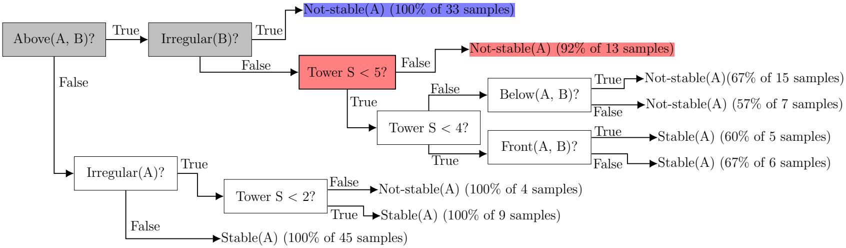
Architecture Components: Inductive Learning



- Input: features and figure labels.
- Decision Tree: induction of rules (constraints, causal laws).
- Output: learned rules.

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).

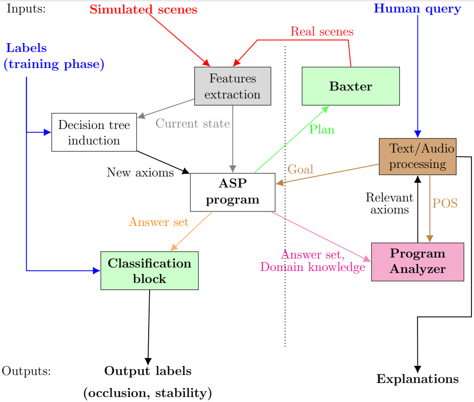
Architecture Components: Inductive Learning



$$\neg stable(A) \leftarrow obj\_rel(above, A, B),\ obj\_surface(B, irregular)$$

$$\neg stable(A) \leftarrow obj\_rel(above, A, B),\ tower\_height(A, N),\ N \geqslant 5$$

Overall Architecture



- Question/request types: (i) describe plan; (ii) why action X? (iii) why not action Y? and (iv) why belief Z?
- Use human input (if available);
- Retrieve relevant information; proof tree belief tracing.

Tracing Example

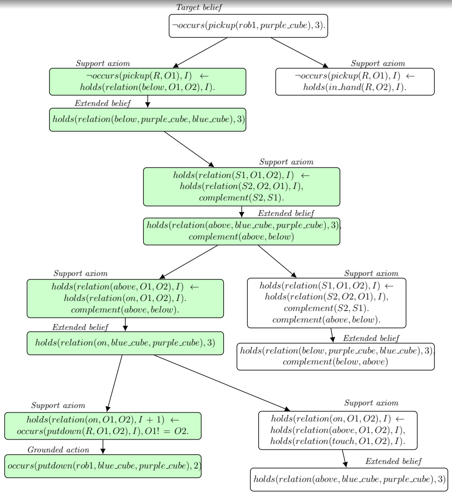
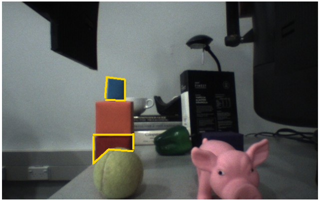


Figure: Example of belief tracing to explain non-selection of an action.

Execution Trace

- **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?"
- ...

Interactive Explanations (Robot Moving Books)

Human: Please describe the executed plan.
Robot: [...] I moved to the $library$. I put down the programming book on $table_1$. I picked up the robotics book from $table_1$. I moved to $office_2$. I gave the large robotics book to the $engineer$. I achieved the goal.
Human: Provide a more detailed description.
Robot: [...] I moved to cell c_2 in the $study$. [...] I moved to cell c_4 in the $library$ and put down the small programming book with a hard cover on the table. I picked up the large robotics book with a soft cover by its spine from the table. [...] I moved to cell c_7 of the $office_2$. I gave the large robotics book to the $engineer$. I achieved the goal.
Human: Why did you move to cell c_4 in $library$?
Robot: I had to put the programming book on $table_1$. I know $table_1$ is in cell c_4 in $library$. I moved to c_4.
Human: Why do you believe the encyclopedia is in $office_2$?
Robot: At time 0, I assumed encyclopedia was in $library$. At time 4, I observed this was false. To restore consistency, I inferred it was in $office_2$."

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

Experimental Results: Learning + Explanations

Precision and recall for learning previously unknown axioms using decision tree induction.

Missing Axioms	Precision	Recall
Strict	69.2%	78.3%
Relaxed	96%	95.1%

(**Real scenes**) Precision and recall of retrieving relevant literals for constructing explanations with and without the learned axioms for reasoning.

	Precision		Recall	
Query Type	**Without**	**With**	**Without**	**With**
Plan description	78.54%	100%	67.52%	100%
Why X?	76.29%	95.25%	66.75%	95.25%
Why not X?	96.61%	96.55%	64.04%	100%
Belief	96.67%	99.02%	95.6%	100%

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 11-14, 2020.

Contributions

- Represent, reason, learn jointly with descriptions and mechanisms. Simplifies design, increases confidence, promotes scalability.
- 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.
- Further explore the interplay between reasoning and learning.

More Information

- Interactive learning for spatial relations, axioms: **IJCAI-18**, **RSS-19** (Best Paper Finalist).
- Theory of explanations, explainable reasoning+learning: **FrontiersAI-19**, **KI-19**, **EUMAS-20**.
- Refinement-based architecture: **NMR-14**, **TRO-15**, **IJCAIwrksp-16**, **JAIR-19**.
- Initial work on non-monotonic logic, hierarchical POMDPs for KRR: **ICDL-12** (Paper of Excellence), **ICAPS-08** (Distinguished Paper), **AIJ-10**, **TRO-13**.