Anchoring interaction through symbolic knowledge

Séverin Lemaignan, Emrah Akin Sisbot, Rachid Alami

Anchoring interaction through symbolic knowledge

[Extended Abstract]

Séverin Lemaignan
CNRS-LAAS, 7 avenue du Colonel Roche
Université de Toulouse, UPS, INSA, ISAE, LAAS
F-31077 Toulouse, France
severin.lemaignan@laas.fr

Akin Sisbot
CNRS-LAAS, 7 avenue du Colonel Roche
Université de Toulouse, UPS, INSA, ISAE, LAAS
F-31077 Toulouse, France
sisbot@laas.fr

Rachid Alami
CNRS-LAAS, 7 avenue du Colonel Roche
Université de Toulouse, UPS, INSA, ISAE, LAAS
F-31077 Toulouse, France
rachid.alami@laas.fr

ABSTRACT

This paper presents how extraction, representation and use of symbolic knowledge from real-world perception and human-robot verbal and non-verbal interaction can actually enable a grounded and shared model of the world that is suitable for later high-level tasks like dialogue understanding, symbolic task planning or reactive supervision. We show how the anchoring process itself fundamentally relies on both the situated and embodied nature of human-robot interactions. We present an implementation, including a specialized symbolic knowledge representation system based on Description Logics, and experiments on several robotic platforms that demonstrate these cognitive capabilities.

Keywords

human-robot interaction, symbolic knowledge anchoring, ontologies, natural language processing

1. GROUNDING HUMAN INTERACTION

A messy table, covered with cardboard boxes, books, video tapes... Thomas is moving and packs everything with the help of Jido, its robot.

" – Jido, give me this", says Thomas, looking at a video tape. The robot smoothly grasps the tape, and hands it to the human.

While this kind of interaction should hopefully sound quite familiar in a forseeable future, our robots are not yet quite up to the task. Neither regarding natural language understanding nor plan-making and manipulation.

To be combined together, those abilities require first an unambiguous and shared representation of concepts (objects, agents, actions...) underlying the interaction. We want besides this representation to remain at the same time close to human understanding to ease interactions, and easy to process and useful for machine.

HRI Pioneers Workshop 2010 Lausanne, Switzerland

Our work focus on these questions: what are the prerequisites for such a human sentence — "Jido, give me this" — to be understood by the robot, correctly interpreted in the spatial context of the interaction, and ultimately transformed into an action?

Numerous work in the field of grounded speech interaction can be mentionned [4, 3] that develop relevant paradigms and demonstrate working integrations on simple robots.

Likewise, several teams have been working on to symbol grounding for robots in real-world environments (computation of geometric relationships from an on-line reconstructed 3D model of the environment and a symbolic framework based on ontologies) [2, 1].

Besides proposing a new integration model for natural language processing with symbolic knowledge repositories, our work extend these previous contribution by focusing on more realistic human-robot interactions: open speech ; complex, dynamic, partially unknown human environments ; fully embodied (with arms, head, ability to move,...) autonomous robots that can perform manipulation. Another original contribution in the field of robotics can be found in its explicit modeling of other agents: each agent is endowed with its own, independent knowledge model in the robot's architecture, containing their beliefs of the world (viewed from the robot, i.e. the robot's beliefs about the beliefs of the agents). We show how leveraging *visual* and *spatial perspective taking* in conjunction with a sound symbolic knowledge representation system enables resolution of natural language interactions and active supervision of the robot.

Lastly, the tools (all open-source) and architecture we present have been deployed and tested on three distinct robotic platforms (including both humanoid robot and service robots), demonstrating the versatility and hardware-

agnosticity of our developments.

2. APPROACH

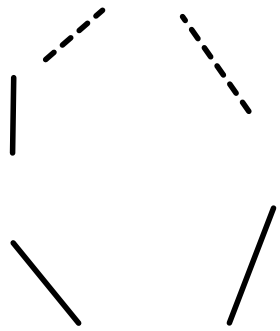


Figure 1: Generic model of cognitive abilities interactions for grounding

Our work introduce three distinct cognitive functions integrated into a full cognitive architecture:

1) Physical environment modeling and *spatial reasoning* are in charge of rebuilding a coherent physical model of the world (figure 2). Once available, the geometric model is used to compute several spatial properties of the scene that actually convert the original sensory data into symbolic beliefs. This includes relative locations of objects, visibility state, gestures like pointing, etc. Assuming that other agents are as well represented in the model, the same computations can be applied to analyze the scene from each agents' point of view (*i.e.* from their *perspectives*).

2) Knowledge representation and management: the robot should build and keep up-to-date a logically sound symbolic model of its beliefs on the world, as well as models for each cognitive agent the robot interact with. Each of these models should be individually consistent, but they do not have to be necessarily globally consistent (for instance, a specific object can be visible for some agent and non-visible for another one). We have built the *ORO* platform that enables continuous storage, querying and reasonning over the pool of facts known by the robot.

Finally *3) dialogue input processing*, including natural language parsing capabilities, disambiguation routines and interactive concept anchoring. Orders, questions, statements (new information) are recognized and separately processed.

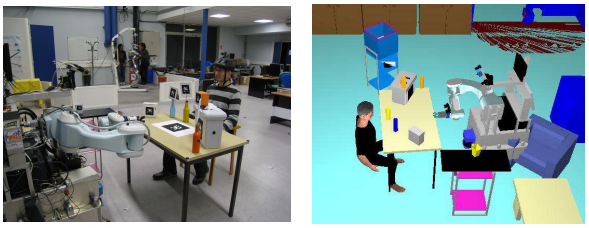
3. EXPERIMENTS

Several experiments have been conducted on different robots.

1. The *Odd One out* task has be conducted on the Rosie robot at Munich university. It uses the *ORO* knowledge base to anchor perception into the robot's knowledge through interaction with the user: the robot picks an unknown object from the table, shows it to the user, and asks about its name and type. The user describes the object (through the categories it belongs to) until a concept known by the robot is given. The learning process starts over again with another unknown object. Once all objects are learned, the robot tells which objects do not belong to a typical breakfast table (*i.e.* objects that are neither food or tableware).

2. The *Spy Game* experiment, conducted at LAAS on the Jido robot, is based on the traditional children game "I Spy". The idea is to discover the object or concept one of the participants is thinking of by asking questions such as: "Is it green? Is it a machine? Is it on your left?", etc. When playing, children exploit their knowledge about the world while categorizing and describing objects through useful discriminants that will allow them to find out the answer as fast as possible while including perspective taking abilities [**?**].

3. The *Moving to London* scenario is a daily life situation: Tom and Jerry are moving to London, so they are packing things in boxes. The scenario takes places in the living-room, where the robot is observing while they move things here and there. To assess the reasoning abilities of the robot, they query it for information (entered through keyboard) or ask it to perform actions.



(a) (b)

Figure 2: Using a dynamic 3D model for situation assessment: (a) Real environment and (b) 3D environment model, viewed in Move3D.

4. REFERENCES

[1] M. Beetz, L. Mösenlechner, and M. Tenorth. CRAM — A Cognitive Robot Abstract Machine for everyday manipulation in human environments. In *International Conference on Intelligent RObots and Systems*, 2010.

[2] M. Daoutis, S. Coradeshi, and A. Loutfi. Grounding commonsense knowledge in intelligent systems. *Journal of Ambient Intelligence and Smart Environments*, pages 311–321, 2009.

[3] G. Kruijff, P. Lison, T. Benjamin, H. Jacobsson, H. Zender, I. Kruijff-Korbayová, and N. Hawes. Situated dialogue processing for human-robot interaction. *Cognitive Systems*, pages 311–364, 2010.

[4] R. Ros, S. Lemaignan, E. A. Sisbot, R. Alami, J. Steinwender, K. Hamann, and F. Warneken. Which one? grounding the referent based on efficient human-robot interaction. In *19th IEEE International Symposium in Robot and Human Interactive Communication*, 2010. To appear.

[5] D. Roy. A computational model of three facets of meaning. In A. G. Manuel de Vega and A. Graesser,

editors, *Symbols and Embodiment - Debates on meaning and cognition*. Oxford University Press, 2008.