Deeper Understanding of Vague Instructions through Simulated Execution

(Extended Abstract)

Mihai Pomarlan
Institute for Artificial Intelligence
University of Bremen
blandc@cs.uni-bremen.de

Daniel Nyga
Institute for Artificial Intelligence
University of Bremen
blandc@cs.uni-bremen.de

Mareike Picklum
Institute for Artificial Intelligence
University of Bremen
blandc@cs.uni-bremen.de

Sebastian Koralewski
Institute for Artificial Intelligence
University of Bremen
seba@cs.uni-bremen.de

Michael Beetz
Institute for Artificial Intelligence
University of Bremen
beetz@cs.uni-bremen.de

ABSTRACT

A commonsense understanding of the physical world will be crucial for the robots of the future as they strive to perform everyday activities and instructions formulated by human users in natural language. One mechanism that is believed to assist human cognition in commonsense reasoning is mental simulation, the envisioning of actions before they are performed. We therefore present a system integrating simulation of robot plans with probabilistic reasoning about natural-language instructions, to create a complete pipeline from instruction to execution to storing and analyzing results of the simulation. This integration allows the robotic system to efficiently infer knowledge about the physical world that would be tedious to specify by hand in a collection of logical statements. Our system will be available online[1] for open use by researchers.

Keywords

Simulation techniques tools and platforms, simulated plan execution, natural language understanding, physics reasoning

1. MOTIVATION AND OVERVIEW

In order to interact with humans in everyday activities, robots must understand and appropriately respond to instructions formulated in human language. However, natural language instructions are often incomplete and rely on implicit, "commonsense" knowledge to interpret. Also, while the AI community has largely moved away from believing symbolic manipulation on its own is capable to produce meaning, the issue of where meaning could come from in robot cognition (what is often referred to as the grounding problem) has no consensus solution.

One approach to the grounding problem has been the simulation theory of cognition [1]. This has been put forth as a hypothesis for where human naive physics knowledge comes from, and has since been adopted by roboticists as a means to ground robot understanding: a robot can understand an action if it can simulate it, and based on the simulation, perform inferences about the action. For example, in [2] a simulation is used to retrieve information about the force required to pick up a rare egg without breaking it. In [3] a system is presented which learns suitable positions to put down an object through repeated simulations. Simulation techniques can also indicate whether an arrangement of objects is stable [4].

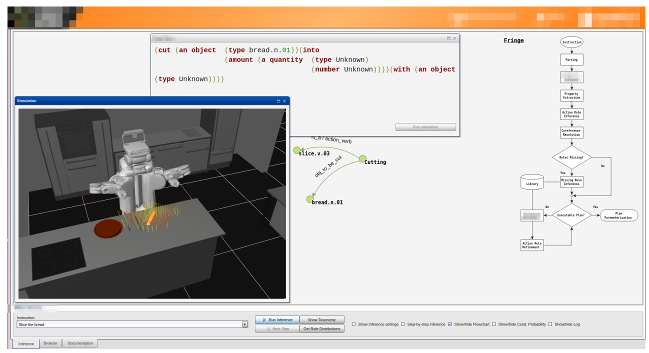

Figure 1: A screenshot of our system, showing a plan call generated from a natural-language instruction and the simulation window.

Motivated by the promise of the simulation theory of cognition, we present a system (available online at `http://prac.informatik.uni-bremen.de/`) that implements a complete pipeline, from parsing instructions given in natural language, to selecting a plan to execute them, to execution, and collection of data from the executed run. The execution of the instructions is performed in a simulated environment. We are also interfacing our system with IAI Bremen's OPENEASE (`http://open-ease.org/`), so that data collected from simulation can be analyzed and queried by a human user as soon as a simulation completes.

Our system aims to provide an environment in which to explore the applications of simulation to robot cognition: what knowledge can be extracted from simulation that is not found in the natural language instructions that triggered it, learning parametrizations for plans, inferring from experience rules of thumb about object behavior (such as which containers can hold which objects), inferring from experience what simulation settings are particularly demanding of robot behavior robustness. Our system is openly available and we will continue adding features to the user interface, such as simulation parametrization and plan library selection.

[1] `http://prac.informatik.uni-bremen.de/`

Appears in: *Proc. of the 16th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2017), S. Das, E. Durfee, K. Larson, M. Winikoff (eds.), May 8–12, 2017, São Paulo, Brazil.*

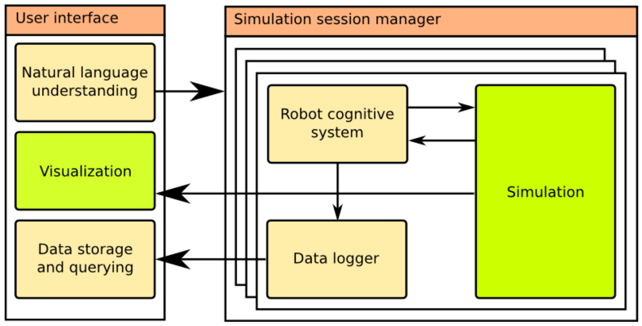

Figure 2: Architecture overview of our system. For reasons of responsiveness, several simulated worlds are running concurrently, waiting for a command from the natural-language understanding component. The simulation produces logs of data about the robot's processes, and a live visualization for the user.

Figure 2 shows an overview of the architecture of our system. On the server side, we use our PRAC system [5] to understand natural-language queries and interpret them into a format containing an action name and role-value pairs for the identified action parameters; some of these parameters can be extracted from the instruction itself, others are inferred probabilistically based on training corpora of what parameters tend to appear together with an action.

The command is then sent to the simulation manager. For reasons of responsiveness (since it takes several seconds to start a simulation), we maintain a pool of simulated worlds available. The manager will select a world appropriate for the command (some commands will be for simulations in a kitchen context, others in a chemical lab, and we plan to add more such worlds in the future), then send the command to the selected world, where the command is further interpreted by the simulated robot's cognitive architecture, converted into an action plan, then executed. The various nodes of the simulated robot– its sensors, controllers, the cognitive component– are all implemented as ROS nodes. A live visualization of the simulation is provided via Robot Web Tools; it allows the user to see the simulated robot and objects, as well as any visualization markers created by the robot plan. Interaction with the simulation through the visualization window is not yet supported. The robot inside the simulation is controlled by the Cognitive Robot Abstract Machine (CRAM) [6], which is a system for enabling the programming and execution of reactive, flexible, taskable robot behaviors.

Another important component is the logger, which records information from the cognitive component of the simulated robot, as well as the simulator's output with regard to object movement. The data from a simulation is stored on our server; we are connecting our system to IAI Bremen's OPENEASE so as to allow users to visualize and query this data after a simulation completes.

2. RELATED WORK

The simulation theory of cognition puts forth the hypothesis that "thinking is simulated interaction with the environment" [7], and can be traced back to older ideas in philosophy (see [7]). More recently, it has been detailed and explored in neuroscience, for example in studying how something like mental simulation may accelerate feedback for motor control [8], language understanding [9], physics knowledge and understanding other minds [1].

In the field of neuroscience, several pieces of research indicate that simulation may be a useful mental process for language understanding [10, 9], physics reasoning and understanding other minds [1], and may be crucial for achieving expert performance [11]. Research has also shown that similar neural regions are active during performance of an activity in the real world, as well as just imagining the activity [12]. It is possible that humans use a "noisy" or approximate physics model [13], possibly non-Newtonian. Also, the quick acquisition of language and motor skills by very young children suggests they are learned in parallel and reinforce each other [14].

The simulation theory of cognition has also been applied to the research and development of artificial agents. Physics simulation was used to tackle a benchmark commonsense reasoning problem, the "Egg Cracking" problem [15], where it proved more scalable and easier to use than previous axiom-based approaches. Other uses have involved training controllers for walking [16] or cutting [17] in simulation before attempting real actions, and obtaining priors for body tracking [18].

There are also criticisms of the simulation theory of cognition [19, 20]. In brief, the arguments are that the veridical, detailed representations simulations require are not compatible with the incomplete knowledge an agent has about their world; simulations are inaccurate idealizations; they have limited usefulness beside prediction. Of course, we take the opposing view that simulation can handle the grounding and frame problems much easier than other approaches, that inaccuracy does not prevent simulation to produce good guesses, and that simulating several scenarios can give an agent the evidence to infer more general principles.

Natural language interpretation by artificial agents is a robustly active research field. As a few examples, probabilistic methods have been explored to infer vague instruction parameters: Markov Logic Networks are used in [5], Conditional Random Fields by [21]. In [22], a system is presented that extracts machine-readable activity knowledge from instructional web-sites aimed at humans.

In terms of architectures for producing and managing robot behavior, we mention several. ROSCo [23] allows human users to specify and share robot behaviors represented as hierarchical state machines. The Cognitive Robot Abstract Machine (CRAM) [6] defines a richer plan language to specify generic plans that are adaptable, taskable, and with failure recovery; we use CRAM for our simulated robot's cognitive component. The system of [24] integrates task and motion planning into a hierarchical behavior generation system. This is developed further in [25], where the planning is performed in belief space. Of particular interest to this paper is the system presented in [21], which also combines natural language, simulation, and robot execution of actions. However, the focus of that research was for a robotic system to learn what actions are appropriate as a response to natural language queries, based on demonstrations from human users in simulated environments. In contrast, we are here interested in what knowledge a robotic agent can learn from simulating its own behavior.

Acknowledgements

This work is supported by the EU FP7 project *RoboHow* (Grant Agreement Number 288533).

REFERENCES

[1] K. Markman, J. A. Suhr, and W. M. P. Klein, *Handbook of Imagination and Mental Simulation*. Psychology Press, 2008.

[2] L. Kunze, M. E. Dolha, E. Guzman, and M. Beetz, "Simulation-based temporal projection of everyday robot object manipulation," in *10th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS)*, Taipei, Taiwan,

May, 2–6 2011.

[3] L. Mösenlechner and M. Beetz, "Parameterizing Actions to have the Appropriate Effects," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, San Francisco, CA, USA, September 25–30 2011.

[4] L. Kunze, M. E. Dolha, and M. Beetz, "Logic Programming with Simulation-based Temporal Projection for Everyday Robot Object Manipulation," in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, San Francisco, CA, USA, September, 25–30 2011, best Student Paper Finalist.

[5] D. Nyga and M. Beetz, "Cloud-based Probabilistic Knowledge Services for Instruction Interpretation," in *International Symposium of Robotics Research (ISRR)*, Sestri Levante (Genoa), Italy, 2015.

[6] M. Beetz, L. Mösenlechner, and M. Tenorth, "CRAM – A Cognitive Robot Abstract Machine for Everyday Manipulation in Human Environments," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Taipei, Taiwan, October 18-22 2010, pp. 1012–1017.

[7] G. Hesslow, "The current status of the simulation theory of cognition," *Brain Research*, vol. 1428, pp. 71–79, 2012.

[8] D. M. Wolpert and Z. Ghahramani, "Computational principles of movement neuroscience," *Nature Neuroscience*, vol. 3, pp. 1212–1217, 2000.

[9] B. Kaup, J. Lüdtke, and C. Maienborn, ""the drawer is still closed": Simulating past and future actions when processing sentences that describe a state," *Brain and Language*, vol. 112, no. 3, pp. 159–166, 2010.

[10] J. Feldman and S. Narayanan, "Embodied meaning in a neural theory of language," *Brain and Language*, vol. 89, no. 2, pp. 385 – 392, 2004, language and MotorIntegration.

[11] S. L. Beilock and I. M. Lyons, "Expertise and the mental simulation of action," 2012.

[12] J. Decety and F. Michel, "Comparative analysis of actual and mental movement times in two graphic tasks," *Brain and cognition*, vol. 11, no. 1, pp. 87–97, 1989.

[13] A. N. Sanborn, V. K. Mansinghka, and T. L. Griffiths, "Reconciling intuitive physics and newtonian mechanics for colliding objects." *Psychological review*, vol. 120, no. 2, p. 411, 2013.

[14] D. R. Bailey, "When push comes to shove: A computational model of the role of motor control in the acquisition of action verbs," Ph.D. dissertation, University of California, Berkeley, 1997.

[15] B. Johnston and M. Williams, "A generic framework for approximate simulation in commonsense reasoning systems," in *The AAAI Spring Symposium on Logical Formalizations of Commonsense Reasoning*, 2007, pp. 71–76.

[16] K. Mombaur and K. Berns, *Modelling, Simulation and Optimization of Bipedal Walking*. Springer, 2013.

[17] P. Long, W. Khalil, and P. Martinet, "Robotic deformable object cutting: From simulation to experimental validation," in *European Workshop on Deformable Object Manipulation (EWDOM)*, 2014.

[18] M. Vondrak, L. Sigal, and O. C. Jenkins, "Dynamical simulation priors for human motion tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 52–65, 2013.

[19] E. Davis and G. Marcus, "The scope and limits of simulation in cognitive models," *CoRR*, vol. abs/1506.04956, 2015.

[20] ——, "The scope and limits of simulation in automated reasoning," *Artif. Intell.*, vol. 233, no. C, pp. 60–72, Apr. 2016.

[21] D. K. Misra, J. Sung, K. Lee, and A. Saxena, "Tell me dave: Context-sensitive grounding of natural language to manipulation instructions," in *Proceedings of Robotics: Science and Systems*, Berkeley, USA, July 2014.

[22] J. Ryu, Y. Jung, K. min Kim, and S. H. Myaeng, "Automatic extraction of human activity knowledge from method-describing web articles," in *Proceedings of the 1st Workshop on Automated Knowledge Base Construction*, 2010, p. 16.

[23] H. Nguyen, M. T. Ciocarlie, K. Hsiao, and C. C. Kemp, "Ros commander (rosco): Behavior creation for home robots." in *ICRA*. IEEE, 2013, pp. 467–474.

[24] L. P. Kaelbling and T. Lozano-Pérez, "Hierarchical task and motion planning in the now," in *IEEE International Conference on Robotics and Automation, ICRA 2011, Shanghai, China, 9-13 May 2011*, 2011, pp. 1470–1477.

[25] ——, "Integrated task and motion planning in belief space," *I. J. Robotics Res.*, vol. 32, no. 9-10, pp. 1194–1227, 2013.