Robotic Agents that Know What They are Doing

Michael Beetz and Daniel Beßler and Gayane Kazhoyan*
{beetz, danielb, kazhoyan}@cs.uni-bremen.de

***Abstract*— Explicit representations of robot knowledge are necessary to achieve competent robotic agents capable of performing variations of tasks. While state-of-the-art knowledge representations exist and are necessary, they are insufficient, as they are designed to abstract away from how actions are executed. We argue that representations that extend to subsymbolic motion and perception level are needed to fill in the gap. We propose a knowledge infrastructure that supports these representations and provides a web-based service for humans and robots to easily access and exchange the knowledge.**

I. Introduction

In 2002 Ronald Brachman gave a seminal talk titled ”Systems That Know What They Are Doing” [1] at the symposium for the cognitive systems program of DARPA – the DARPATech 2002. In this talk he was pushing the idea of cognitive computer systems where he defined the latter in the following way: ”A truly cognitive system would be able to learn from its experience – as well as by being instructed – and perform better on day two than it did on day one. It would be able to explain what it was doing and why it was doing it. It would be reflective enough to know when it was heading down a blind alley or when it needed to ask for information that it simply couldn't get to by further reasoning. And using these capabilities, a cognitive system would be robust in the face of surprises. It would be able to cope much more maturely with unanticipated circumstances than any current machine can.” The DARPA cognitive systems program was groundbreaking, creating the Cognitive Assistant that Learns and Organizes (CALO) system [2] and the Siri agent as one of its commercial offsprings. The program was focusing on and limited to software agents that assisted knowledge workers in cooperative office work such as reading emails, organizing meetings, project management etc. These systems had huge success for symbolic information processing tasks. Eventually, these developments led to the creation of the Watson system [3].

While these major advances were achieved in the symbolic and text-based information processing world, nothing comparable has yet happened in the area of autonomous robots. The need for having a similar revolution in the autonomous robotics domain is more and more widely accepted in the community, perhaps most explicitly spelled out in the article by Gill Pratt on the ”Cambrian explosion” in robotics [4].

We believe that the big barriers in realizing robotic agents that know what they are doing are the difficulty in modeling action and behavior representations, and explicitly stating commonsense and naive physics knowledge that relates motions to their effects. We require cognitive agents to be able to answer queries such as ”How did the robot grasp an object?”, ”Where did it stand during the grasping action?”, etc. We take the ability to answer such queries as the criteria of being cognitive in the sense of Brachman's definition. Being able to answer these queries requires a formal representation of actions, their effects, the motions they generate, the intentions and beliefs of the robot, as well as the causal and teleological relations between these concepts.

In order to obtain this knowledge, we propose that the robotic agents record their experiences of interacting with the world and annotate them with automatically built up symbolic representations of this information. The experiences are subsymbolic recordings of the robot's sensor and motion streams, annotated with a symbolic narrative. The latter is generated by, in a way, making the robot talk to itself about the decisions, reasoning and perceptions that it performs during action execution, and time-synchronizing the resulting story with the subsymbolic data. This enables robots to perform semantic retrieval of their experiences. This means that we can learn definitions of symbolic concepts that can be formulated items of the narrative in terms of subsymbolic data. For example, we can learn an answer to the query: ”Where should I stand to pick up an object successfully?”, and the answer will be a geometric entity.

Our research agenda is to learn grounded models of manipulation actions that are general in a way that they can be transferred to and exchanged between different robots, and to collect these generalized representations into a knowledge base for a web-based knowledge service openEASE (Fig.1).

Fig. 1. openEASE web interface (`http://open-ease.org`)

II. Knowledge Service Queries

The recorded experiences can be queried using a Prolog-based language which expands to temporal logics reasoning. The heterogeneous experience data is hooked into the

*The authors are with the Collaborative Research Centre Everyday Activities Science and Engineering (EASE), University of Bremen, Germany.

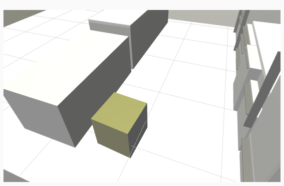
temporal reasoner via procedural attachments to symbols in the terminological knowledge base. The knowledge base appears to be purely symbolic to the end user while internally symbols may be computed on demand from heterogeneous data sources such as sensor data.

Relevant entities such as objects and actions are terminologically described in the knowledge base, and can be queried using partial entity descriptions, similar to SQL queries. This enables, for example, to ask questions such as "Where did the robot stand when it performed a put-down action?":

Query ?

```
entity(Tsk, [an, action,
    [type, putting-down-an-object],
    [performed-by, [a, robot,
        [part, [an object,
            [type, mobile-base],
            [name, Base]]]]]]),
occurs(Tsk, [_,End]),
entity(Base, [an, object,
  [pose, Pose, during, End]]),
show(cube(mobile_base), [pose(Pose)]).
```

This query reads as follows: "Find an action *Tsk* that was performed by a robot with mobile base *Base*, which ended at the time instant *End*. Furthermore, compute the pose of the mobile base at time instant *End*." The answer to this question is the set of bindings for variables appearing in the query accompanied by the scene visualization:



```
Tsk   = log:'PuttingDownAnObject_0',
Base  = robot:'MobileBase_pr2',
Pose  = pose([TX,TY,TZ], [RW,RX,RY,RZ]),
End   = 1396512613.0.
```

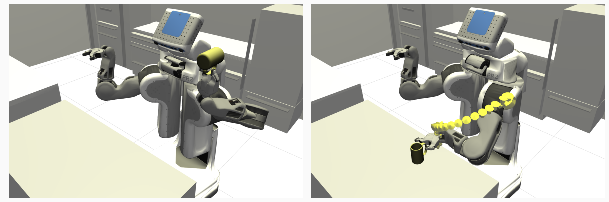
! Answer

Various aspects of actions covered by NEEMs can be accessed through entity queries. For example, questions such as "How did the robot move its end effector while performing put-down actions?" can easily be stated as entity query:

Query ?

```
entity(Tsk, [an, action,
    [type, putting-down-an-object],
    [subaction, [an, action,
        [type, release-grasp],
        [body-part, Part],
        [grasp-spec, Spec]]]]),
occurs(Tsk, [Begin,End]),
entity(Part, [an, object,
    [type, gripper],
    [trajectory, Tr, during, [Begin,End]]]),
show(Tr).
```

Which reads as: "Find an action *Tsk* during which an object was put down, with a sub-action during which the end effector *Part* was releasing the grasp according to the grasp specification *Spec*, and which occured during the time interval *[Begin,End]*. Furthermore, compute the trajectory *Tr* of the end effector *Part* and visualize it."



```
Tsk   = log:'PuttingDownAnObject_0',
Part  = robot:'GripperLeft_pr2',
Spec  = objects:'GraspSpecification_0',
Tr    = comp:'Trajectory_GripperLeft_pr2_0',
Begin = 1396512603.0,
End   = 1396512613.0.
```

! Answer

A. Generalizing Knowledge

Symbolically annotated experiences are comprehensive recordings of actions from which learning problems can be derived by filtering the experiences according to criteria specified in entity queries. The queries need to be written such that the data is relayed to context and task. Action parameters over which to generalize can be determined through clustering. The base pose for putting down objects on a table, for example, can be learned from experiences that capture successful put-down actions:

Query ?

```
findall(Tsk,
  entity(Tsk, [an, action,
        [type, putting-down-an-object],
        [success, true],
        [target-location, [an, object,
          [type, table]]]],
      Tsks).
```

III. Conclusion

Having semantic representations of actions available at such detailed level, including subsymbolic motions and data extracted from images, we expect all the decision problems done by the robot to be eventually automatically turned into supervised learning problems through issuing the respective queries to the knowledge base.

Acknowledgments

This work was supported by the DFG collaborative research center #1320 "Everyday Activity Science and Engineering (EASE)".

References

[1] R. J. Brachman, "Systems that know what they're doing," *IEEE Intelligent Systems*, vol. 17, no. 6, pp. 67–71, Nov 2002.

[2] G. Tur, A. Stolcke, L. Voss, S. Peters, D. Hakkani-Tur, J. Dowding, B. Favre, R. Fernandez, M. Frampton, M. Frandsen, C. Frederickson, M. Graciarena, D. Kintzing, K. Leveque, S. Mason, J. Niekrasz, M. Purver, K. Riedhammer, E. Shriberg, J. Tien, D. Vergyri, and F. Yang, "The CALO Meeting Assistant System," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 6, pp. 1601–1611, Aug 2010.

[3] D. Ferrucci, E. Brown, J. Chu-Carroll, J. Fan, D. Gondek, A. A. Kalyanpur, A. Lally, J. W. Murdock, E. Nyberg, J. Prager, *et al.*, "Building Watson: An overview of the DeepQA project," *AI magazine*, vol. 31, no. 3, pp. 59–79, 2010.

[4] G. A. Pratt, "Is a Cambrian Explosion Coming for Robotics?" *Journal of Economic Perspectives*, vol. 29, no. 3, pp. 51–60, September 2015. [Online]. Available: http://www.aeaweb.org/articles?id=10.1257/jep.29.3.51