Towards Evaluating Human Robot Dialog based Affordance Learning and the Challenges

Amit Kumar Pandey and Coline Le Dantec and Rodolphe Gelin

Abstract Perception of Affordances shapes how we interact with environment. A novel way to learn affordances at symbolic level is through human-robot natural dialog. The challenge is how to evaluate such learned knowledge. As a first step this paper studies how do humans associate affordances and objects in a given scenario and can it be used as ground truth to evaluate the content learned by the robot.

1 Introduction

Reasoning about affordance, [1], [2] - what an object offers as action possibilities - is important to shape our day-to-day interaction with the environment, and a central organizing construct for action, [3], and embodied cognition, [4].

This paper focuses on the idea that the same object can afford different action possibilities when associated with different other objects and depends upon the context, [5], [6]. In Robotics, affordances have been perceived through vision, [7], manipulation trials, [5], etc. In [8] we have demonstrated a novel way of learning affordance through *natural dialog based interaction*. This is inspired by the studies suggesting that language and verbs reflect some characteristics of action organization and evoke simulations, which in turn activate functional affordances, [9]. In [8], a *Nao* robot in an office tabletop scenario through 32 interaction instances with a user over multiple days, has learned two types of associations of affordance: *Object-Object association* for a particular affordance and *Affordance-Object association* for different affordances. In this paper, we will discuss some preliminary ideas in the direction of evaluating such dialog based affordance learning.

2 Objective

Affordance learning through social and dialog based interaction is mostly symbolic and potentially at different levels of abstraction. Therefore, the main **challenge** is to investigate, "*what is the ground truth of such affordance learning?*" In this direction we are exploring *whether the humans collective response gives a generalized answer?* For this, a set of analyses is under progress. Two of them presented here are with the **objective** *to understand how people associate affordances to a given scenario and compare with what the robot has learned*. People have been asked to:

1. Group an unordered list of objects based the imagined action possibilities.
2. List the potential activities based on a given scenario and list of objects.

Aldebaran SoftBank Group, Paris, France.
akpandey@aldebaran.com; cledantec@aldebaran.com; rgelin@aldebaran.com
Funded by *Romeo2* project (http://www.projetromeo.com), BPIFrance in the framework of PSPC

3 Preliminary Results

People responded through an online questionnaire. The scenario, context and objects were similar to the situation in which the robot had earlier learned affordances in [8]. A total of 23 responses received, mostly from our robotics R&D department.

Object-Object binding based on affordance: Some of the object-object bindings, which the robot learned in [8] based on its experience, were: *(Keyboard, Monitor), (Pen, Paper), (Monitor, Mouse), (Monitor, Mobile), (Headphone, Laptop), (Notebook, Pen), (Monitor, Pen), (Bottle, Lunch Box)*. Figure 1 shows the humans responses for point 1 (of sec. 2) for grouping an unordered list of objects (*coffee cup - computer - books - paper - pens - smartphone*). This shows some similarity with how the robot has paired objects based on learned affordance, such as (*Paper, Pen*).

Object based Affordance learning: For the (*Monitor, Keyboard*) objects pair, some of the object based affordances, which the robot learned in [8], based on its experience, were: *Program, Work, Talk, Check-email*. Figure 2 shows the humans responses for point 2 (of sec. 2) about potential activities in a similar scenario. This also shows some similarity with what the robot has learned as affordances.

4 Open Questions and Challenges

Despite of showing some agreements between robot's learned knowledge and humans' responses, at the same time these preliminary analyses are raising the question, "*Should Human response be the benchmark test*?" "*Does there exist a real ground truth*?" Because, from people's response we got non-consistent answers, lists and ordering of object-object and object-affordance associations.

Also the high percentage of *To Work* for *(Monitor, Keyboard)* affordance, make us to think about the underlying structure of affordances. Perhaps most people first thought about *To Work* as a high level affordance and then further elaborate with more specific affordances like *To Check mails*, *To Program*, etc. at lower levels. Or perhaps "*Work*" might be a totally different affordance for people so as the *To Check mails*, *To Program*. Therefore, it also points to an interesting direction to investigate: *whether a hierarchy of affordance exit or affordance concepts are flat?*

We also found that the majority of the responses were confined to the action possibilities in the office environment and domain. This raises another interesting dimension to *explore the role or influence of context and scenario with how the robot should perceive functional affordance.*

Answering all these will have high relevance for designing and development of coherent theoretical and functional frameworks of affordance perception and learning for robots in human centered environment.

5 Conclusion and Work in Progress

This preliminary analysis suggests that it is not trivial to find a ground truth for evaluating dialog based symbolic level of learned functional affordances by the robot. There is a great need of investigation to address the question "*how functional affordances are perceived and represented at symbolic level and how to evaluate them*". We are also investigating in the direction of *what to mainly evaluate in such dialog based affordance learning?* Should it be "*What has been learned*", "*How fast and how it has been learned*", "*Why it has been learned*" or their combinations.

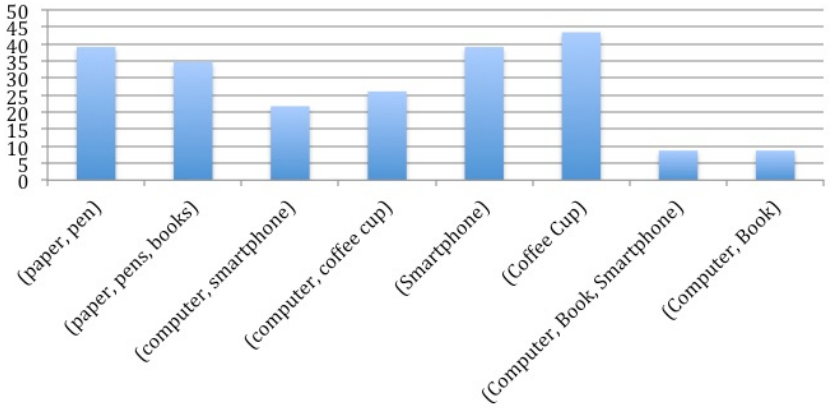

Fig. 1: Grouping of objects based on how humans perceived various action-possibilities. Y-axis is the percentage of people grouped those objects.

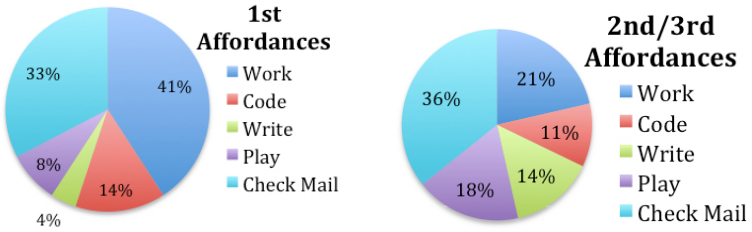

Fig. 2: Participants' suggested action possibility based Primary and secondary affordances, extracted for *computer* in an office scenario.

References

1. J. J. Gibson, "The theory of affordances," in *The Ecological Approach To Visual Perception*. Psychology Press, Sep. 1986, pp. 127–143.
2. D. Norman, *The psychology of everyday things*. Basic Books, 1988.
3. A. Clark, "An embodied cognitive science?" *Trends in cognitive sciences*, vol. 3, no. 9, pp. 345–351, 1999.
4. F. Garbarini and M. Adenzato, "At the root of embodied cognition: Cognitive science meets neurophysiology," *Brain and cognition*, vol. 56, no. 1, pp. 100–106, 2004.
5. B. Moldovan, P. Moreno, M. van Otterlo, J. Santos-Victor, and L. De Raedt, "Learning relational affordance models for robots in multi-object manipulation tasks," in *IEEE International Conference on Robotics and Automation (ICRA),*. , 2012, pp. 4373–4378.
6. H. A.Pieropan, C.H.Ek, "Recognizing object affordances in terms of spatio-temporal object-object relationships," in *International Conference on Humanoid Robots, Madrid, Spain*, 2014.
7. H. S. Koppula, R. Gupta, and A. Saxena, "Learning human activities and object affordances from rgb-d videos," *Int. J. Rob. Res.*, vol. 32, no. 8, pp. 951–970, Jul. 2013.
8. A. K. Pandey, R. Gelin, "Human robot interaction can boost robot's affordance learning: A proof of concept," in *International Conference on Advanced Robotics (ICAR)*, pp. 642–648, 2015.
9. A. M. Borghi, "Action language comprehension, affordances and goals." in Yan Coello, Angela Bartolo (eds), Language and action in cognitive neuroscience. Contemporary topics in cognitive neuroscience series. Psychology press, pp. 125-143.