Semantic Representation of Objects and Function

Vasanth Sarathy[1] and Matthias Scheutz[2]

I. INTRODUCTION & MOTIVATION

Natural human activities involve using and manipulating objects, often several at a time, simultaneously and continuously. For example, changing lanes while driving requires using a steering wheel and pedals, all while observing the road and making sure that lanes are clear. Not only are we recognizing these objects, but we know what to do with them (i.e., we can perceive object affordances). We use our imagination and invoke mental simulations to construct variations of objects and actions to learn these affordances. Learning how to use objects is a highly desirable skill for robots, as well. Unfortunately, although robots are proficient at recognizing object features, they are less-skilled at recognizing what can be done with these objects.

In this paper, we sketch a novel approach based on Dempster-Shafer (DS) theory [1] and uncertain logic for inferring object affordances using a mental simulation framework. As part of this effort, we are in the process of developing a computational model for affordance that can represent complicated activities and can account for the dynamic and continuous nature of real-world scenarios.

II. BACKGROUND

Gibson introduced the concept of affordance to represent the relationship between an agent and its environment [2]. In cognitive science, Barsalou et al. expanded this work and attached causality to function and affordance [3]. In cognitive robotics, Montesano et al. have developed statistically-inspired causal models of affordance using Bayesian Networks (BN) to formalize the relationship between object features, actions and effects [4].

Despite these efforts, affordance-learning faces many challenges that have not been overcome in the previous work including: integrating multi-modal cues in real-time (natural language, vision, gesture, mental simulation); accounting for situational information and how an object may be used in a given context; representing complex affordance relationships in dynamic environments consisting of a sequence of situations; inferring causal, non-causal and counterfactual relationships from highly-limited data; and representing uncertainty in knowledge and beliefs.

[1]Vasanth Sarathy is a Graduate Student in Computer Science and Cognitive Science, Human-Robot Interaction Laboratory, Tufts University, 200 Boston Ave., Medford, MA, USA vasanth.sarathy@tufts.edu

[2]Matthias Scheutz is a Professor in the Department of Computer Science, Human-Robot Interaction Laboratory, Tufts University, Medford, MA USA matthias.scheutz@tufts.edu

III. OUR APPROACH

Much like the previous work, our efforts begin with Gibson's definition of affordance as a relationship between an agent and its environment. However, we diverge from the previous work in our representational and computational approach to modeling affordance.

We propose a model, illustrated in Fig. 1, in which an object's affordance (F) and the perceived feature of the object (O) depend on the context (C). The perceived feature of the object (O) also depends on affordance (F).

We use Dempster-Shafer (DS) theory for inferring affordance (F) from object features (O) in contexts (C), and conversely, for generating objects (O) during mental simulation from affordances (F) in contexts (C). DS theory is an uncertainty processing framework often interpreted as a generalization of the Bayesian framework. A Bayesian approach for inferring $P(F|O,C)$ by way of $P(O|F,C)$, $P(F|C)$, and $P(C)$ is not practical because we do not have a probability distribution for all the affordances for an object. Instead, we use rules of the form $o \wedge c \implies_{[\alpha,\beta]} f$ that captures the affordance behind an object in particular contexts, where $[\alpha,\beta]$ is a confidence interval contained in $[0,1]$, which can be specified for the rules independently. These rules are very versatile and we can employ DS-theoretic modus ponens to make uncertain deductive and abductive inferences which cannot be made in a mere Bayesian framework.

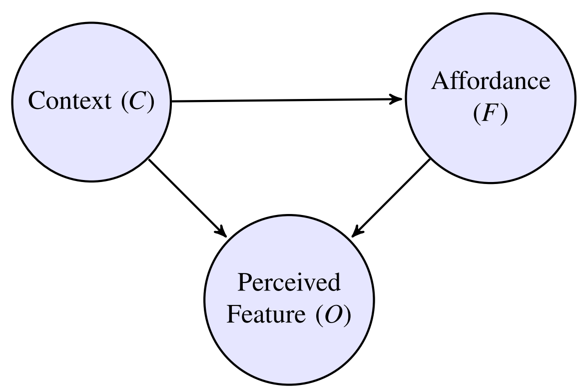

Fig. 1. Object's affordances (F) depends on the context (C). The perceived features of the object (O) depends on the context (C) it is being perceived in and on the affordance (F). For example, in the context of hammering a nail (C), the elongated end of a hammer is perceived as being a "handle" (O) as a result of it being "grabable" (F). Other parts of the hammer may be perceived as serving as a "handle" in different contexts. For example, we may grab the head of the hammer and use the elongated end as an extender to reach something otherwise out of reach. In this context, the head of the hammer is perceived as being the handle O, with the affordance of "grabable" (F) in the context of reaching (C).

IV. HANDLING UNCERTAINTY

Consider the example of robot learning how to tighten a screw. We would like for the robot to understand this task from an intuitive standpoint such that even in the absence of a screwdriver, it can reason through alternatives and find another substitute. While demonstrating the task, a human teacher explains the she is "holding the handle," "inserting the flat-head end into the screw," and the "turning the screw clockwise to tighten." In this example, the semantic representation of the screwdriver may be described as a series of predicates:

$Shape(elongate)$
$EndFeature(\text{“}flathead\text{”}, flat)$
$BodyFeature(\text{“}handle\text{”}, textured)$
$EndFeature(\text{“}top\text{”}, round)$

The robot may infer a number of rules about this task from the human, as well. One rule may be: that given a goal to tighten a screw, and given a flat-head end feature, the screw can be tightened by inserting the flat-head into the screw and turning clockwise. This rule can then be represented in DS-theoretic uncertain logic a follows:

$$r^0_{[\alpha_{R_0}, \beta_{R_0}]} = (EndFeature(\text{“}flathead\text{”}, flat)) \wedge Goal(self, tighten, screw) \implies Tightenable(screw, action(insert, flathead, screw), action(turn, clockwise))$$

The robot's past knowledge (which it may have acquired from other unrelated demonstrations, mental simulations and real-world exploration) may include an additional set of facts about knives and coins as shown in Table 1:

TABLE I
KNOWLEDGE BASE

Knives	$Shape(elongate)$ $EndFeature(\text{“}sharpedge\text{”}, flat)$ $BodyFeature(\text{“}handle\text{”}, smooth)$ $EndFeature(\text{“}top\text{”}, square)$
Coins	$Shape(disk)$ $EndFeature(\text{“}thickness\text{”}, flat)$ $BodyFeature(\text{“}faces\text{”}, heads, tails)$

By applying DS-theoretic logic inference rules on this knowledge base, the robot can deduce how to tighten a screw by using a knife or coin instead of a screwdriver. Although, the rules in this example are relatively simple, we do envision scenarios that involve more complicated rules, or bundles of rules for a more complex object representation.

V. EVENTS

Real-world scenarios are not static and generally involve a dynamic and continuous sequence of actions and events. For example, the act of tightening a screw involves several steps of finding a screwdriver, grabbing it, and then tightening the screw. This continuous set of actions can be broken up into discrete chunks by slicing it at times when events occur. For example, occurrences such as touching the screwdriver, or the screwdriver touching the screw, or the screw being fully tightened are events. We propose extending our DS-theory based representational framework to account for these types of events by building joinder rules defining how they demarcate the boundary between actions. For example, a touch event of a screwdriver touching the screw might signal the start of the process of turning the screw.

Under our proposed framework, object features have their own uncertainty interval. Inference in uncertain logic adds the possibility of tracking and propagating uncertainties that may arise on premises and/or rules. Consequently, we can perform modus ponens, leading us to deducing new facts (with some uncertainty) about the objects themselves.

VI. CONCLUSION

As part of their interview process, many modern technology companies show prospective candidates an object they have never seen before and ask them to describe what they think is the object's function. The purpose of the question is to test the candidate and probe their intellect to identify candidates with strong mental representations of affordance. Clever answers are often rewarded and stand as an example of human creativity. The ultimate goal of our research is to endow robots with the ability to find creative ways to use and manipulate objects and their environment, especially when there is minimal and uncertain information. Such abilities will be highly desirable in open-world scenarios such as search-and-rescue missions.

In this extended abstract, we took the first steps towards our goal and sketched a novel approach based on Dempster-Shafer (DS) theory for inferring object affordances. By example, we provided an overview for how our framework can handle uncertainties and be extended to include the continuous and dynamic nature of real-world situations. We have developed a computational framework for implementing a DS-theoretic approach in a different context, and we are currently working to implement those algorithms and architecture for modeling object affordance. We believe that this, much richer level of affordance representation is needed to allow robots to be adaptable to novel open-world scenarios.

REFERENCES

[1] Shafer, G., A Mathematical Theory of Evidence, Princeton University Press, 1976.
[2] J. J. Gibson, The Ecological Approach to Visual Perception, Boston: Houghton Mifflin, 1979.
[3] Barsalou, L. W., Sloman, S. A, and Chaigneau, S. E., The HIPE theory of function. In L. Carlson and E. van der Zee (Eds.), Representing functional features for language and space: Insights from perception, categorization and development. Oxford, England: Oxford University Press.
[4] Montesano, L., Lopes, M.; Bernardino, A., Santos-Victor, J., Modeling affordances using Bayesian networks, Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on , vol., no., pp.4102,4107, Oct. 29 2007-Nov. 2 2007.
[5] Williams, T.; Briggs, G.; Oosterveld, B.; Scheutz, M., Going Beyond Literal Command-Based Instructions: Extending Robotic Natural Language Interaction Capabilities, AAAI Conference on Artificial Intelligence, North America, Feb. 2015.