Cognitive Affordance Representations in Uncertain Logic

Vasanth Sarathy
Tufts University, Medford, MA 02155, USA
vasanth.sarathy@tufts.edu

Matthias Scheutz
Tufts University, Medford, MA 02155, USA
matthias.scheutz@tufts.edu

Abstract

The concept of "affordance" represents the relationship between human perceivers and their environment. Affordance perception, representation, and inference are central to commonsense reasoning, tool-use and creative problem-solving in artificial agents. Existing approaches fail to provide flexibility with which to reason about affordances in the open world, where they are influenced by changing context, social norms, historical precedence, and uncertainty. We develop a formal rules-based logical representational format coupled with an uncertainty-processing framework to reason about cognitive affordances in a more general manner than shown in the existing literature. Our framework allows agents to make deductive and abductive inferences about functional and social affordances, collectively and dynamically, thereby allowing the agent to adapt to changing conditions. We demonstrate our approach with an example, and show that an agent can successfully reason through situations that involve a tight interplay between various social and functional norms.

Introduction

Natural human activities involve using and manipulating objects around us and reasoning about our environment. Sometimes these activities involve standard reasoning tasks, like changing lanes while driving, which requires using a steering wheel and pedals, all while observing the road and making sure that lanes are clear. Other times, these activities involve more creative reasoning tasks like solving puzzles and finding novel uses for objects. When performing these activities not only are we recognizing these objects in our environment, 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 infer these affordances. We then use these affordances to reason about the task at hand. Learning how to use objects is a highly desirable skill for artificial agents, 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 present a novel approach based on Dempster-Shafer (DS) theory (Shafer 1976) and "uncertain logic" for inferring object affordances. 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. We will demonstrate our approach with an example involving contextual and commonsense reasoning.

Background

James Gibson (1979) introduced the concept of "affordance" to represent the relationship between the agent and its environment. Past work in formalizing this relationship has largely focused on modeling affordance using either statistical formalisms or ontology-based approaches. For example, Montesano et al. have developed statistically inspired causal models of affordance using Bayesian Networks to formalize the relationship between object features, actions and effects (Montesano et al. 2007). Varadarajan et al. (Varadarajan 2015) developed a detailed knowledge-ontology based on conceptual, functional and part properties of objects, and then used a combination of detection and query matching algorithms to pinpoint the affordances for objects.

Despite these efforts, affordance representation faces many challenges that have not been overcome in the previous work. These approaches fail to provide flexibility with which to reason about affordances in the open world, where they are influenced by changing context, social norms, historical precedence, and uncertainty. For example, these current approaches cannot reason that coffee mugs afford grasping and drinking, while also affording serving as a paperweight or cupholder, or depending on the context, as family heirloom not meant to be used at all.

Representing Cognitive Affordances

We propose a novel model and formal rules-based logical representational format for cognitive affordances, in which an object's affordance (A) and its perceived features (F) depend on the context (C). We use Dempster-Shafer (DS) theory for inferring affordance (A) from object features (F) in contexts (C). DS theory is an uncertainty processing framework often interpreted as a generalization of the Bayesian framework.

The proposed cognitive affordance model consists of four parts: (1) a set of perceivable object features (F), (2) a set of context states (C), (3) a set of object affordances (A), and (4) a set of affordance rules (R) from object features and context states to affordances taking the overall form:

$r :\equiv f \wedge c \implies_{[\alpha,\beta]} a$

with $f \in F$, $c \in C$, $a \in A$, $r \in R$, $[\alpha, \beta] \subset [0, 1]$. Here, the confidence interval $[\alpha, \beta]$ is intended to capture the uncertainty associated with the affordance rule r such that if $\alpha = \beta = 1$ the rule is logically true, while $\alpha = 0$ and $\beta = 1$ assign maximum uncertainty to the rule. Rules can then be applied for a given feature percept f in given context c to obtain the implied affordance a under uncertainty about f, c, and the extent to which they imply the presence of a. We have previously shown that these types of rules are very versatile and that we can employ DS-theoretic modus ponens to make uncertain deductive and abductive inferences (Williams et al. 2015). Most critically, these rules allow us to address representational challenges with mere Bayesian models such as inferring $P(A|F,C)$ by way of $P(F|A,C)$, $P(A|C)$, and $P(C)$ when we often have no practical way to obtain the necessary probability distributions for all the affordances for an object. We will next review the basics of Dempster Shafer theory to be able to further discuss our intended use of these rules.

Dempster-Shafer Theory Preliminaries

A set of elementary events of interest is called *Frame of Discernment* (FoD). The FoD is a finite set of mutually exclusive events $\Theta = \theta_1, ..., \theta_N$. The power set of Θ is denoted by $2^\Theta = A : A \subseteq \Theta$ (Shafer 1976).

Each set $A \subseteq 2^\Theta$ has a certain weight, or *mass* associated with it. A *Basic Belief Assignment* (BBA) is a mapping $m_\Theta(\cdot) : 2^\Theta \to [0,1]$ such that $\sum_{A \subseteq \Theta} m_\Theta(A) = 1$ and $m_\Theta(\emptyset) = 0$. The BBA measures the support assigned to the propositions $A \subseteq \Theta$ only. The subsets of A with non-zero mass are referred to as *focal elements* and comprise the set $\mathcal{F}_\Theta$. The triple $\mathcal{E} = \{\Theta, \mathcal{F}_\Theta, m_\Theta(\cdot)\}$ is called the *Body of Evidence* (BoE). For ease of reading, we sometimes omit $\mathcal{F}_\Theta$ when referencing the BoE.

Given a BoE $\{\Theta, \mathcal{F}_\Theta, m_\Theta(\cdot)\}$, the *belief* for a set of hypotheses A is $Bel(A) = \sum_{B \subseteq A} m_\Theta(B)$. This belief function captures the total support that can be committed to A without also committing it to the complement A^c of A. The *plausibility* of A is $Pl(A) = 1 - Bel(A^c)$. Thus, $Pl(A)$ corresponds to the total belief that does not contradict A. The $uncertainty$ interval of A is $[Bel(A), Pl(A)]$, which contains the true probability $P(A)$. In the limit case with no uncertainty, we get $Pl(A) = Bel(A) = P(A)$.

Logical inference with uncertainty can be performed using DS-theoretic Modus Ponens (denoted $\odot$) (Tang et al. 2012). We will use the DS-theoretic AND (denoted $\otimes$) to combine BoEs on different FoDs (Tang et al. 2012), and Yager's rule of combination (denoted $\bigcap$) to combine BoEs on the same FoD (Yager 1987). We choose to use Tang's models of Modus Ponens and AND over other proposed models because those models do not allow uncertainty to be multiplicatively combined. Similarly, Yager's rule of combination is chosen because it allows uncertainty to be pooled in the universal set, and due to the counter-intuitive results produced by Dempster's rule of combination, as discussed in (Zadeh 1979).

Algorithm 1 getAffordance($\{\Theta_F, m_f\}$,$\{\Theta_C, m_c\}$,R)

1: $\{\Theta_F, m_f\}$: BoE of candidate perceptual features
2: $\{\Theta_C, m_c\}$: BoE of relevant contextual items
3: R: Currently applicable rules
4: $S = \emptyset$
5: **for all** $r \in R$ **do**
6: $\quad S = S \cup \{(m_f \otimes m_c) \odot m_{r=fc \to a}\}$
7: **end for**
8: $G = group(S)$
9: $\psi = \emptyset$
10: **for all** group $g_a \in G$ **do**
11: $\quad \psi = \psi \cup \{\bigcap_{j=0}^{|g_a|} g_{a_j}\}$
12: **end for**
13: **return** ψ

For two logical formulae ϕ_1 (with $Bel(\phi_1) = \alpha_1$ and $Pl(\phi_1) = \beta_1$) and ϕ_2 (with $Bel(\phi_2) = \alpha_2$ and $Pl(\phi_2) = \beta_2$, applying logical AND yields $\phi_1 \otimes \phi_2 = \phi_3$ with $Bel(\phi_3) = \alpha_1 * \alpha_2$ and $Pl(\phi_3) = \beta_1 * \beta_2$.

For logical formulae ϕ_1 (with Bel$(\phi_1) = \alpha_1$ and $Pl(\phi_1) = \beta_1$) and $\phi_{\phi_1 \to \phi_2}$ (with $Bel(\phi_{\phi_1 \to \phi_2}) = \alpha_R$ and $Pl(\phi_{\phi_1 \to \phi_2}) = \beta_R$, the corresponding model of Modus Ponens is $\phi_1 \odot \phi_{\phi_1 \to \phi_2} = \phi_2$ with $Bel(\phi_2) = \alpha_1 * \alpha_R$ and $Pl(\phi_2) = 1 - ((1 - Pl(\beta_1)) * (1 - Pl(\beta_R)))$

Moreover, we will use the "ambiguity measure" λ defined in (Nunez et al. 2013) to be able to compare uncertainties associated with formulas ϕ and their respective confidence intervals $[\alpha, \beta]$:

$$\lambda(\alpha, \beta) = 1 + \frac{\beta}{\gamma} log_2 \frac{\beta}{\gamma} + \frac{1-\alpha}{\gamma} log_2 \frac{1-\alpha}{\gamma}$$

$$\text{where } \gamma = 1 + \beta - \alpha.$$

Here, ϕ is deemed more ambiguous as $\lambda(\alpha, \beta) \to 0$.

Inferring Affordances with Uncertain Logic

To infer cognitive affordances, we propose to start with the first prototype inference algorithm shown in Algorithm 1 and refine it to tailor it specifically to a cognitive affordance model.

The algorithm takes three parameters: (1) a BoE of candidate perceptions $\{\Theta_F, m_f\}$ is provided by the low-level vision system, (2) a BoE of relevant contextual items $\{\Theta_C, m_c\}$ provided by a knowledge base or some other part of the integrated system that can provide context information, and (3) a table of cognitive affordance rules R.

The inference algorithm then examines each rule $r_{fc \to a} \in R$ (line 5), and performs DS-theoretic AND and Uncertain Modus Ponens to obtain m_a from $m_{fc \to a}$ and m_{fc} (line 6).

Note that since we allow multiple affordance rules to be considered, multiple affordances may be produced. Multiple rules may produce the same affordances for various reasons, possibly at different levels of belief or disbelief. However, we seek to return the set of *unique* affordances implied by a set of perceptions f.

After considering all applicable affordance rules, we group affordances that have the same content but different mass assignments (line 8), and use Yager's rule of combination (line 11) to fuse each group of identical affordances.

Finally, we can use the ambiguity measure λ to determine whether an inferred affordance should be realized and acted upon. For example, we could check the ambiguity of each affordance $a \in \psi$ on its uncertainty interval $[\alpha_i, \beta_i]$: if $\lambda(\alpha_i, \beta_i) \leq \Lambda(c)$ (where $\Lambda(c)$ is an ambiguity threshold, possibly depending on context c), we do not have enough information to confidently accept the set of inferred affordances and can thus not confidently use the affordances to guide action. However, even in this case, it might be possible to pass on the most likely candidates to other cognitive systems. Conversely, if $\lambda(\alpha_i, \beta_i) > \Lambda(c)$, then we take the inferred affordance to be certain enough to use it for further processing.

Example: Using and Handing Over Objects

We will now present an evaluation of the proposed representation and algorithm, and demonstrate the capabilities facilitated by this approach with an example of using and handing over objects in a kitchen. Handing over objects "properly" is an important skill for helper robotic agents. When performing a handover, the robot will need to reason about the normative and contextual factors that influence the propriety of a handover.

We will represent the agent's knowledge about being a kitchen-helper with 9 rules along with their respective uncertainty intervals (Figure 1). We have chosen uncertainty intervals in such a way that the more specific the rule, the more certainty and higher degree of belief the agent has about that particular rule.

Consider a robotic agent helper receiving instructions from a human, Julia. Suppose Julia says to the robot: *"Bring me something clean I can use to cut this tomato."* The robotic agent, $X = self$, parses this request from Julia and assigns its own task-context and determines the types of affordances it is interested in exploiting in the kitchen-environment. The agent is confident that it is in the kitchen context and it further determines, with a high degree of certainty, that it is in the context of handing over an object in the kitchen, and assigns context masses as follows:

$domain(self, kitchen)$: $m_{c_{1,1}} = 1.0$
$task(self, give, O)$: $m_{c_{2,1}} = 0.95$
$task(self, use, O)$: $m_{c_{2,1}} = 0.05$

At this point, the agent can review its environment and examine each object more closely. Based on the set of rules, it knows to look specifically for certain visual percepts stated in the rules, such as $near()$, $sharpEdge()$, and so on.

Let us assume that the agent spots a knife on the counter. Upon reviewing the physical features of the knife, the agent determines masses for the relevant visual percepts cited in the rules (e.g., $hasSharpEdge(knife) = 0.95$). The agent examines each rule $r^i_{[\alpha_i,\beta_i]}$ (per Algorithm 1) to obtain m_a from $m_{fc\rightarrow a}$ and m_{fc}.

Commonsense Physical Rules:
$r^1_{[0.8,1]} := hasSharpEdge(O) \wedge$
$domain(X, kitchen) \implies$
$cutWith(X, O)$

General Social Rules:
$r^2_{[0.95,0.95]} := \neg inUse(O, H) \wedge$
$domain(X, kitchen) \implies$
$graspable(X, O, holdPart(O, P))$
$r^3_{[0.95,0.95]} := \neg inUse(O, H) \wedge$
$domain(X, kitchen) \implies$
$graspable(X, O, funcPart(O, P))$

General Object Grasp Rules:
$r^4_{[0.55,0.95]} := near(O, G, holdPart(O, P)) \wedge$
$domain(X, kitchen) \implies$
$graspable(X, O, holdPart(O, P))$
$r^5_{[0.55,0.95]} := \neg near(O, G, holdPart(O, P1)) \wedge$
$near(O, G, funcPart(O, P2)) \wedge$
$domain(X, kitchen) \implies$
$graspable(X, O, funcPart(O, P2))$

Task-based Social Rules:
$r^6_{[0.8,0.9]} := near(O, G, holdPart(O, P)) \wedge$
$task(X, use, O) \wedge$
$domain(X, kitchen) \implies$
$graspable(X, O, holdPart(O, P))$
$r^7_{[0.8,0.9]} := near(O, G, funcPart(O, P)) \wedge$
$task(X, give, O) \wedge$
$domain(X, kitchen) \implies$
$graspable(X, O, funcPart(O, P))$
$r^8_{[0.95,0.95]} := near(O, G, holdPart(O, P)) \wedge$
$\neg dirty(O)$
$task(X, use, O) \wedge$
$domain(X, kitchen) \implies$
$graspable(X, O, holdPart(O, P))$
$r^9_{[0.95,0.95]} := near(O, G, funcPart(O, P)) \wedge$
$\neg dirty(O)$
$task(X, give, O) \wedge$
$domain(X, kitchen) \implies$
$graspable(X, O, funcPart(O, P))$

Figure 1: Cognitive Affordance Rules for a Kitchen-Helper Agent.

For example, consider rule r^1:

$$r^1_{[0.8,1]} := hasSharpEdge(O) \wedge domain(X, kitchen) \implies cutWith(X, O)$$

The agent will apply perceptual and contextual information as follows, to determine the affordance implied by the rule:

$$\frac{r^1_{[0.8,1]}(m_r = 0.8) := hasSharpEdge(knife)(m_f = 0.95) \wedge domain(self, kitchen)(m_c = 1.0) \implies}{cutWith(self, knife)(m_a = (m_f \otimes m_c) \odot m_r = 0.76)}$$

The uncertainty interval for the rule can then be computed as $[0.76, 1]$.The agent will then perform this analysis for each of the other rules in the set to determine uncertainty intervals for the implied affordances.

To be able to generate a set of unique affordances, a, implied by feature, f, after considering all applicable affordance rules, we thus group affordances that have the same semantic content but different mass assignments (e.g., $graspable()$) and use Yagers rule of combination to fuse each group of identical intentions, adding the resulting fused intention to set ψ.

Thus, we can generate a list of unique affordances available to the agent at the current moment in time, when it has seen the knife:

Available affordances (Upon seeing the knife), ψ
$cutWith(knife)[0.76, 1], \lambda = 0.29$
$graspable(self, knife,$ $holdPart(knife, handle))[0.96, 0.99], \lambda = 0.78$
$graspable(self, knife,$ $funcPart(knife, blade))[0.98, 0.99], \lambda = 0.88$

The agent might decide that because there is a high degree of certainty that the object under consideration has a $cutWith$ affordance, it will choose to grasp it and then select to grasp it at the blade (as opposed to the handle), to accomplish a socially apropriate handover.

Discussion

We use very simple predicate-style descriptions for percepts, context and affordances. Note, we have neither used a systematic ontology to represent these components, nor have we elected a formal language. However, the proposed algorithm and inference mechanism are sufficiently general-purpose to work with any suitable formalism. Whatever formalism is selected, we argue for an explicit representation of context because it can help constrain the set of affordance rules as well as guide what the agent must attend to in its perception of the world.

Thus far, we have not discussed, expressly, the origin of the cognitive affordance rules and how an agent might generate or learn new rules, because this is not the focus of the paper. Nevertheless, we expect these rules can be learned in a number of different ways from explicit demonstration and instruction (Cantrell et al. 2012), from observation through reinforcement learning (RL) techniques (Boularias, Bagnell, and Stentz 2015) or from exploration, and using multiple different modalities including vision, natural language and haptic information.

The proposed computational model is general and showcases the potential of an affordance-based uncertain logic reasoning process. Reasoning about cognitive affordances in a more general way, as outlined in this paper, has the potential to assist in commonsense and creative reasoning (e.g., by finding rules in the agent's extended knowledge-base that have similar affordances but are associated with different contexts) as well as in sense-making (e.g., by reasoning about a situation looking at affordances of objects in a scene collectively).

Conclusion and Future Work

In this paper, we proposed a novel framework and algorithm based on Dempster-Shafer (DS) theory for inferring object affordances. We demonstrated how the proposed framework can handle uncertainties and be extended to include the continuous and dynamic nature of real-world situations, which we believe is needed to allow artificial agents to be adaptable to novel open-world scenarios.

The proposed framework has shown some potential, but we will still need to address various challenges in computational complexity as well as ontology and definition for context, and how the agent might move from one context to another. We will look into a framework for learning various affordance rules, and in doing so, we will need to incorporate action formalisms to allow a more dynamic reasoning process. Finally, we will be demonstrating the proposed framework on robotic systems and grounding our representation to work with embedded cognitive systems.

Acknowledgements

This research was funded in part by ONR grant #N00014-14-1-0149.

References

Boularias, A.; Bagnell, J. A.; and Stentz, A. 2015. Learning to Manipulate Unknown Objects in Clutter by Reinforcement. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence Learning*, 1336–1342.

Cantrell, R.; Talamadupula, K.; Schermerhorn, P.; Benton, J.; Kambhampati, S.; and Scheutz, M. 2012. Tell Me When and Why to Do It!: Run-time Planner Model Updates via Natural Language Instruction. In *Proceedings of the 2012 Human-Robot Interaction Conference*.

Gibson, J. J. 1979. *The ecological approach to visual perception*, volume 39.

Montesano, L.; Lop, M.; Bernardino, A.; and Santos-Victor, J. 2007. Modeling affordances using Bayesian networks. In *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 4102–4107.

Nunez, R. C.; Dabarera, R.; Scheutz, M.; Briggs, G.; Bueno, O.; Premaratne, K.; and Murthi, M. N. 2013. DS-based uncertain implication rules for inference and fusion applications. *Information Fusion (FUSION), 2013 16th International Conference on* 1934–1941.

Shafer, G. 1976. *A Mathematical Theory of Evidence*. Princeton University Press.

Tang, Y.; Hang, C. W.; Parsons, S.; and Singh, M. 2012. Towards argumentation with symbolic dempster-shafer evidence. *Frontiers in Artificial Intelligence and Applications* 245(1):462–469.

Varadarajan, K. 2015. Topological mapping for robot navigation using affordance features. *2015 6th International Conference on Automation, Robotics and Applications (ICARA). Proceedings*.

Williams, T.; Briggs, G.; Oosterveld, B.; and Scheutz, M. 2015. Going Beyond Literal Command-Based Instructions : Extending Robotic Natural Language Interaction Capabilities. In *AAAI*, 1387–1393.

Yager, R. R. 1987. On the dempster-shafer framework and new combination rules. *Information Sciences* 41(2):93–137.

Zadeh, L. A. 1979. *On the Validity of Dempster's Rule of Combination of Evidence*. Electronics Research Laboratory, University of California.