Relational affordance learning for task-dependent robot grasping

Laura Antanas[1], Anton Dries[1], Plinio Moreno[2], Luc De Raedt[1]

[1] Department of Computer Science, Katholieke Universiteit Leuven, Belgium
[2] Institute for Systems and Robotics, IST, University of Lisboa, Portugal

Abstract. Robot grasping depends on the specific manipulation scenario: the object, its properties, task and grasp constraints. Object-task affordances facilitate semantic reasoning about pre-grasp configurations with respect to the intended tasks, favouring good grasps. We employ probabilistic rule learning to recover such object-task affordances for task-dependent grasping from realistic video data.

1 Introduction and related work

Robot grasping skills are essential for acting in dynamic environments. Objects can be grasped in different ways depending on the specific manipulation scenario: the object, its properties, task and grasp constraints. Inspired by the definition of object affordances – which refers to the properties of an object to allow actions to be performed on it by a human or other entity, we investigate the benefits of object-task affordances for task-dependent grasping in a kitchen environment. Our earlier work on task-dependent grasping [2] shows that, when combined with probabilistic reasoning and object/task ontologies, they facilitate compact grasping models which generalize over object/task categories in a natural way, while showing robustness to uncertainty and missing information. Here we propose, as key contribution, a statistical relational learning approach to learn object affordances for task-dependent grasping.

Let us consider the scenario in Fig. 1. A mobile robot with grasping capabilities must grasp a bottle from the shelf and place it on the table. The environment constraints (e.g. narrow spaces) and task constraints (e.g. the most stable pre-grasp gripper pose for grasping the bottle) present a difficult problem which can be solved using semantic reasoning. If we consider the top, middle and bottom as semantic parts of the bottle, the best part to grasp it is from the middle, given that it needs to be placed on the table upright and the top is partially obstructed by the shelf above. Given such semantic object parts (or pre-grasps), object properties, and the intended task, we can learn probabilistic grasp-related rules for our kitchen scenario, e.g., that a bottle affords pick and placing on a surface by grasping it from the middle. The resulting task-dependent affordances give the robot the capability to semantically reason about the best pre-grasp and thus, help the grasp planner. Our experiments show that we can learn reliable relational affordances from realistic and uncertain video data.

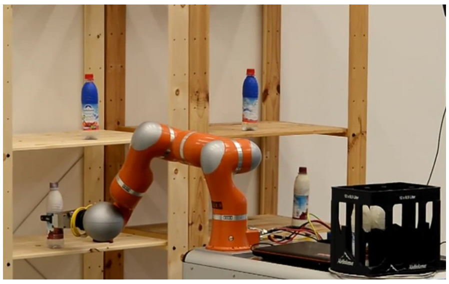

Fig. 1: Manipulation scenario: grasp the bottle from the shelf and place it upright on the table.

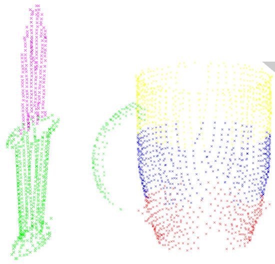

Fig. 2: Semantic parts for knife and cup: yellow-top, blue-middle, red-bottom, green-handle, and magenta-usable area.

Affordances have been considered before in robot manipulation. While in [12] the authors employ estimated visual-based latent affordances, the work in [4] reasons about grasp selection by modeling affordance relations between objects, actions and effects using either a fully probabilistic setting or a rule-based ontology. In contrast, we employ a SRL approach to learn object affordances which generalize over similar object parts and object/task categories. Closely related is the semantic grasping pipeline in [5]. It employs a semantic affordance map which relates gripper approach directions to particular tasks. We exploit additional world knowledge in form of ontologies. This allows us to experiment with a wide range of object categories. Further, to compute plans comprising sequences of actions and to solve complex manipulation tasks, [1] combines symbolic reasoning and learning from demonstrations. In [9] meaningful symbolic relational representations are used to solve sequential manipulation tasks in a goal-directed manner via active relational reinforcement learning. Related work for generalizing over doors and handles using SRL has been proposed in [10].

However, none of these frameworks solves the problem of learning affordances for semantic task-dependent grasping. Relational affordance models for robots have been learned in a multi-object manipulation task context [11]. Differently, we propose learning pre-grasp configurations using task-category affordances. Our approach features semantic generalization and can tackle unknown objects.

2 Problem description and representation

Each scene contains one object and the task to be executed. Its semantic visual description consists of the task, object parts, category, pose, and containment together with their probabilities. In a kitchen scenario, the perception algorithm (see [2]) can segment objects, distinguish between upright and sideways poses and label each part with one of the labels: top, middle, bottom, handle or usable area. This reduces the search space for robot grasp generation, prediction and planning. The object category can be obtained using an object classifier.

The scene is represented as a set of relational visual observations. For the scenario in Fig. 1 they are encoded using probabilistic facts, such as 1.0 :: `object(o)`, stating that an object `o` is observed with probability 1.0. The observation `object(o)` is a logical atom and `object/1` is a predicate symbol of arity 1. The object identifier `o` is a constant and represents a ground term. Terms can also be variables when denoted in uppercase. Ground atoms do not contain variables and represent particular relations. True ground atoms are also called facts. Relational visual observations for our scenario are illustrated in Example 1 (left). We consider the task given, e.g., 1.0 :: $\texttt{task}(\texttt{o}, \texttt{t1}, \texttt{pickPlaceOn})$.

Example 1. Relational representation for our scenario in Fig. 1:

```
object(o).
0.8::category(o,bottle).
0.5::pose(o,upright).
0.5::part(o,p1,top).
0.9::part(o,p2,middle).
0.5::part(o,p3,bottom).
1.0::task(o,t1,pickPlaceOn).
1.0::task(o,t2,pickPlaceInUpsidedown).
```

```
...
1.0::affords(o,t1).
0.0::affords(o,t2).
1.0::impossible(o,t2).
0.0::impossible(o,t1).
0.9::grasp(o,t1,p2).
0.1::grasp(o,t1,p1).
0.0::grasp(o,t1,p3).
```

Next, we define object-task affordances as the tasks afforded by the objects in our robot grasping setup, considering the manipulation capabilities of a gripper mounted on a robotic arm. Fig. 3 illustrates 46 affordances in the form of a table. They allow us to relate object-task concepts as inspired by *AfNet: The Affordance Network*. We consider 11 object categories: {*pan, pot, cup, glass, bowl, bottle, can, hammer, knife, screwdriver, cooking_tool*} and 7 tasks: {*pass, pourOut, pourIn, pickPlaceInUpright, pickPlaceInUpsidedown, pickPlaceInSideways, pickPlaceOn*}. By looking at the table, we can obtain possible object-task affordance pairs. They can be encoded as rules, e.g. $\texttt{affords}(\texttt{X}, \texttt{T}) \leftarrow \texttt{bottle}(\texttt{X}), \texttt{task}(\texttt{T}, \texttt{pass})$ states that a bottle `X` affords the task `T` of passing.

We can make abstraction of fine-grained object categories by plugging in an object category ontology. The super-categories are defined based on the object functionality, and are represented by: {*kitchenContainer, dish, openContainer, canister, container, tool, object*}. The super-category *dish* subsumes the categories *bowl*, *glass* and *cup*. Similarly, tasks can be grouped in super-tasks such as: {*pickPlaceIn, pickPlace, pour, task*}. *PickPlaceInUpsidedown* refers to picking and placing the object inside a shelf in the upside-down pose, *pickPlaceOn* to place the object on a surface in the same pose. Ontologies can be translated into deterministic logical rules which are used by our learner. For example, the rule $\texttt{supercateg}(\texttt{X}, \texttt{container}) \leftarrow \texttt{category}(\texttt{X}, \texttt{bottle})$ states that "any bottle is a container".We can then generally state that any container affords the task of pouring, i.e., $\texttt{affords}(\texttt{X}, \texttt{T}) \leftarrow \texttt{container}(\texttt{X}), \texttt{pour}(\texttt{T})$. However, this is not always true, as pouring liquid in a canister is an almost impossible task, even for a human. We encode such a constraint as $\texttt{impossible}(\texttt{X}, \texttt{T}) \leftarrow \texttt{canister}(\texttt{X}), \texttt{pourIn}(\texttt{T})$, which states that a canister does not afford the task of pouring in. A first goal of this work is to improve robot grasping by learning relational object-task affordances and constrains from data.

Further, depending on the object properties, its parts and task, the object should be grasped in different ways. To reason about good pre-grasp configurations given the intended task, we use the semantic object parts. Similar to

affordances task/object			container							tool			
			open container					canister					
			dish			kitchen							
			cup	glass	bowl	pan	pot	bottle	can	hammer	knife	screwdr	cooking
pass			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
pour		in	✓	✓	✓	✓	✓	-	-	-	-	-	-
		out	✓	✓	✓	-	-	✓	✓	-	-	-	-
p&p	in	upright	✓	✓	✓	✓	✓	✓	✓	-	-	-	-
		upsidedown	✓	✓	✓	-	-	-	-	-	-	-	-
		sideways	-	-	-	-	-	-	-	✓	✓	✓	✓
	on		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Fig. 3: Object-task affordances are marked by ✓, constraints by −.

object categories, pre-grasps can be associated to specific tasks. Each task activates grasping affordances according to associations between object categories, object parts and gripper poses. Besides object category-task associations, the second goal of our work is learning object part-category-task relations. While the first are general pair-wise affordances, the second are grasp-related triplets.

Using the visual observations introduced in Example 1 one can learn several affordances: the bottle affords pick and placing on (a surface), it should be grasped from the middle (from the bottom the gripper might hit the shelf below, from the top the shelf just above), but also constraints: the bottle cannot be placed upsidedown. In order to do so, we also assume to have labeled affordance examples: object category-task relations specified via the target predicate `affords`/2, object category-task constraints via `impossible`/2, and object part-category-task affordances indicated by the target predicate `grasp`/3. Ground target predicates or learning examples are illustrated for our scenario in Example 1 (right). Each example is a fact labeled with a target probability. Target atom 0.9 :: `grasp`(`o`, `t1`, `p2`) states that the bottle `o` can be grasped by the middle with probability 0.9. The resulting set of probabilistic ground facts from all the scenarios considered represent the input for our probabilistic rule learner.

3 Approach

Our examples are facts that are probabilistically entailed by the theory. Thus, we proceed with learning from entailment. We use ProbFOIL+ [6] which generalizes FOIL, mFOIL [8], ProbFOIL [7] and learns probabilistic rules from probabilistic ground facts given as input. The output is a probabilistic classifier in the form of a set of generalized rules that returns a probabilistic target atom. ProbFOIL+ directly generalizes the mFOIL rule learner. It follows a typical sequential covering approach where the outer loop of the algorithm starts from an empty set of clauses and repeatedly adds clauses to the hypothesis until no more improvements are observed with respect to some global scoring function (e.g. accuracy). Each clause is learned in a greedy manner by performing beam search using m-estimate as a local scoring function. What sets ProbFOIL apart from mFOIL is its support for probabilistic data by generalizing the concepts of true/false positive/negative to a probabilistic context. In addition, it performs parameter learning such that it returns rules that express probabilistic relationships.

	CT affordances			CT constraints		PCT affordances	
Dataset	S_{SYN}	S_{REAL}^{perf}	S_{REAL}	S_{SYN}	S_{REAL}	S_{SYN}	S_{REAL}
Rules learned	38 (46)	34 (27)	31 (32)	35 (39)	41 (52)	25 (53)	20 (20)
Accuracy %	86 (89)	99 (99)	88 (90)	79 (92)	85 (90)	83 (86)	85 (87)

Table 1: Number of learned rules and accuracy using fine-grained categories and tasks. The results in brackets consider in addition the pose as input.

Except the target predicate, we provide ProbFOIL+ with a description of the refinement operator in terms of mode declarations (e.g., $\texttt{task}(+, +, \texttt{c})$ which indicates that the first two arguments should be variables that already exist in the clause, and the third argument is to be replaced by a constant). ProbFOIL+ then proceeds by iteratively extending the clause with one literal, pruning the least promising candidates at each step, until no more improvement can be made.

4 Experiments

We experiment on task-dependent robotic grasping datasets for kitchen-related scenarios introduced in [2]. The synthetic dataset denoted S_{SYN} considers flawless detection of objects from 3D meshes. It contains 41 objects belonging to all categories in our ontology and 102 grasping scenarios. The object pose, its parts and object containment are manually labeled, while the object category is estimated via the classifier in [2]. The other 3 datasets are obtained with the ORCA simulator [3]. They contain 25 objects belonging to all categories, except pot and cooking tool, and 134 grasping scenarios. We assume all containers empty. The object part, pose and category recognition modules are external to ORCA. Each object is placed on top of a table. We obtain i) S_{REAL} – object pose, category and its parts are estimated after the point cloud completion. It may have missing parts, when they are occluded or not detected, or extra parts according to the limitations of the detection algorithm; and ii) S_{REAL}^{perf} – we give the groudtruth object category instead of the detected one in S_{REAL} as an upper-bound reference. Our goal is to investigate if we can recover affordances from labeled data and to evaluate how good their rules are. We report accuracy and learned rules for the 3 datasets in Table 1.

Object category-task (CT) affordances. Input features are the task, object category and pose, while $\texttt{affords}/2$ is the target to be learned. This gives a dataset of 714 examples for S_{SYN}, and 882 examples for S_{REAL}. A learned rule for S_{SYN} is $0.78 :: \texttt{affords}(\texttt{A}, \texttt{B}) \leftarrow \texttt{category}(\texttt{A}, \texttt{cup}), \texttt{task}(\texttt{A}, \texttt{B}, \texttt{pourIn})$. We obtain 38 fine-grained rules out of 44 possible (cf. Table 3; our dataset did not contain positive targets pot-pourIn and pan-pourIn). The other depend also on the object initial pose and its containment. When we include them, accuracy increases. Next, using super-categories, we can summarize the set of fine-grained affordances with 15 rules, while keeping the same accuracy. Learned supercategory-task affordances are more general (cup, glass, bowl are replaced by dish, bottle and can by canister, and hammer, screwdriver, cooking tool and knife by tool). For the realistic datasets we can recover 34 fine-grained affordances out of 39 possible (cf. Table 3 without pot and cooking tool) for S_{REAL}^{perf} and 31 for S_{REAL}. We obtain 2 affordances for pot, but none for *pickPlaceInUpsidedown*.

This was due to object misclassifications. By using super-categories, we can replace the set of fine-grained affordances with 7 generalized rules.

Object part-category-task (PCT) affordances. We consider as input information the task, object category, parts and pose, since the part from which to grasp an object for a given task depends on the pose as well; `grasp`/3 is the target to be learned. This gives us a dataset of 2093 examples for S_{SYN} and 2674 for S_{REAL}. Experiments using Probfoil+ give us a grasping model of 53 affordance rules for S_{SYN} and 20 rules for S_{REAL} (when pose is considered). By introducing supercategories the grasp-based models are generalized from 53 rules to 18 and from 20 to 17, respectively, while keeping a close accuracy. A part-category-task affordance learned from S_{SYN} is $0.8 :: \mathtt{grasp(A,B,C)} \leftarrow \mathtt{part(A,C,usable\_area)}, \mathtt{supercategory(A,tool)}, \mathtt{task(A,B,pass)}$.

The use of relational learning techniques for task-dependent grasping contrasts with current approaches that usually learn direct mappings from visual perceptions to object grasping points. Differently, we learn rule-based affordances that generalize over similar object parts and object/task categories and can be used to semantically reason in task-dependent robot grasping. Our experiments show that we can learn reliable relational affordances from realistic video data.

References

1. N. Abdo, H. Kretzschmar, and C. Stachniss. From Low-Level Trajectory Demonstrations to Symbolic Actions for Planning. In *ICAPS Workshop on Combining Task and Motion Planning for Real-World Applications*, pages 1–8, 2012.
2. L. Antanas, P. Moreno, M. Neumann, R. Pimentel de Figueiredo, K. Kersting, J. Santos-Victor, and L. De Raedt. High-level reasoning and low-level learning for grasping: A probabilistic logic pipeline. *CoRR*, abs/1411.1108, 2014.
3. H. Baltzakis. Orca simulator.
4. C. Barck-Holst, M. Ralph, F. Holmar, and D. Kragic. Learning grasping affordance using probabilistic and ontological approaches. In *ICRA*, pages 1–6, 2009.
5. H. Dang and P. K. Allen. Semantic grasping: Planning robotic grasps functionally suitable for an object manipulation task. In *IEEE/RSJ ICIRS*, pages 1311–1317, 2012.
6. L. De Raedt, A. Dries, I. Thon, G. Van Den Broeck, and M. Verbeke. Inducing probabilistic relational rules from probabilistic examples. In *IJCAI*, pages 1835–1843. AAAI Press, 2015.
7. L. De Raedt and I. Thon. *Probabilistic Rule Learning*, pages 47–58. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.
8. S. Džeroski. Handling imperfect data in inductive logic programming. In *SCAI*, pages 111–125, 1993.
9. J. Kulick, M. Toussaint, T. Lang, and M. Lopes. Active learning for teaching a robot grounded relational symbols. In *IJCAI*, pages 1451–1457. AAAI Press, 2013.
10. B. Moldovan, L. Antanas, and M. Hoffmann. Opening doors: An initial SRL approach. In *LNCS*, volume 7842, pages 178–192, 2013.
11. 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 *ICRA*, pages 4373 –4378, 2012.
12. J. Sweeney and R. A. Grupen. A model of shared grasp affordances from demonstration. In *Humanoids*, pages 27–35, 2007.