Inferring Robot Actions from Verbal Commands Using Shallow Semantic Parsing

Alexander Sutherland, Suna Bensch, Thomas Hellström

Department of Computing Science, Umeå University, Sweden

Abstract

Robust natural language understanding is central for the development of robots conversing with humans. In this paper we focus on interpretation of commands given to a robot by a human. The taken approach uses shallow semantic parsing output to infer both the robot action and its parameters. Results show an accuracy of 88% for inference of action alone, and 68% for combined inference of an action and its parameters. At the end of the paper we present an approach on how to relate inferred parameters of robot actions to objects in the real world. To this end, we propose a structural representation of an object that is similar to the dependency representation of the objects' verbal description.

1 Introduction

Natural language is one of the most efficient means of communication for humans and has also been extensively addressed in human-robot interaction research [Bugmann et al.2005, Scheutz et al.2007].

While robust speech recognition is a major unsolved problem in natural language processing (NLP), challenges also remain in other areas of NLP, such as syntactic, semantic or discourse analysis. We abstract from these issues, each of which is still of formidable complexity, and focus on inferring robot actions from natural language utterances, in particular imperative commands (e.g. "Fetch the cup from the kitchen").

We propose a method (published in [Sutherland et al.2015]) to create mappings from sentences to expected robot actions. The approach builds on the hypothesis that an expected robot action can be inferred from shallow parsing output, i.e. collection of frame names and their respective semantic roles. In a learning phase, labeled sentences are semantically parsed using the commonly available Semafor system [Das et al.2010]. The generated frame names and their semantic roles are used to create mappings to expected robot actions and their respective parameters.

We also propose a method [Sutherland2016] of relating the inferred parameters of robot actions to objects in the real world. For this, we propose a structural representation of an object that is similar to the dependency representation of the objects' verbal description.

2 Shallow semantic parsing

Shallow semantic parsing, also called semantic-role-labeling, is the task of finding semantic roles for a given sentence.

Semantic roles describe general relations between predicates and its arguments in a sentence. For example, in a sentence like "Mary gave the ball to Peter", "gave" is the predicate (also referred to as frame name or frames), "Mary" represents the semantic role *donor*, "the ball" represents the semantic role *theme*, and "Peter" represents the semantic role *recipient*. In the present work we use the Semafor system for extraction of frames and semantic roles from all sentences used in the experiments.

3 Inference of actions and parameters

In this section we present the inference of expected robot actions and its parameters.

We propose a method by which the expected robot actions for a verbally uttered command to a robot can be learned, such that the robot automatically can determine what to do when hearing a new sentence. The robot learns how to infer action and parameters from a set of labeled example sentences. The example sentences used in this paper were manually generated. Each sentence was labeled with one of n_A robot actions $a_1, ..., a_{n_A}$ and m_a associated parameters $p_1, ..., p_{m_a}$ (see Table 1).

i	a_i	p_1	p_2	Expected function
1	BRING	object	recipient	*fetches object*
2	TELL	message	recipient	*relays a message*
3	COLLECT	object	source	*gathers objects*
4	MOVE	location		*moves to location*
5	PUT	object	location	*places an object*

Table 1: Pre-programmed robot actions a_i with associated parameters p_1, p_2.

In the learning phase, each sentence in a training data set comprising N sentences was presented to the Semafor system, that outputs frames and its semantic roles. If, for one predicate, several frames were generated, the primary frame, was selected. For our entire data set, $n_F = 21$ distinct primary frames $f_1, ..., f_{n_F}$, were generated. Conditional probabilities for how these entities relate to expected actions and associated parameters are estimated and used to construct the necessary inference mechanisms.

A Bayes classifier is used to infer the expected action a_E for a sentence with a primary frame name f_E and semantic roles $r_i, i = 1, ..., n_R$. Inference of robot action and its parameters for a given sentence is expressed as pseudo-code in Algorithm 1 (for details see [Sutherland et al.2015]).

Algorithm 1 Infer expected action a_E and associated parameters p_i^E for an input sentence s.

return a_E and $p_i^E, ..., p_{m_{a_E}}^E$
inputs:
s : **sentence to be analyzed**
A : **set of training sentences labeled with action** a **and parameters** $p_1, ...p_{m_a}$
$f_E \leftarrow$ **the primary frame of** s
$r_1^E, ..., r_{n_R}^E \leftarrow$ **semantic roles for** s
$B \leftarrow$ **the subset of** A **with** f_E **as primary frame**
$a_E \leftarrow$ **the most common action** a **in** B

for $i = 1$ **to** m_{a_E} **do**
    find the index opt **for which** $p_i \sim r_{opt}$ **in most sentences in** B
    $p_i^E \leftarrow r_{opt}^E$
end for

4 Results

The developed algorithm for inference of robot actions and its parameters was implemented and evaluated using a data set of 94 manually labeled imperative sentences. Evaluation was done by leave-one-out cross-validation.

In order for a robot to be able act correctly on an uttered sentence, both robot action and parameters have to be correctly inferred. The average accuracy for all sentences for inference of robot action was 88%, and for combined inference of robot action and parameters 68%.

5 Future work

In order for a robot to be able to act on parameter values inferred by the method described above, the parameters need to be matched with real world entities. For this, we propose a common structural representation of semantic roles and visual input perceived by the robot.

We abstract from the visual perception module and assume the robot can detect objects in the real world. Semantic roles commonly contain noun phrases that can be described as a set of dependencies as specified by Stanford typed dependencies manual [Marneffe et al.2008], commonly known as a dependency parse tree (DPT). Dependencies can be described as triples of the form $E(V_1, V_2)$ where E is the dependency relationship between some governing word V_1 and some dependent word V_2. In [Sutherland2016] specifically designed rules create a so-called *visual dependency tree* (VDPT) of a recognized object o, similar to a dependency tree of a verbal utterance describing o.

A VDPT t is created roughly as follows: If an object o (e.g. a card) is detected by the vision system, the root node of t is labeled with a label for o (e.g. card). The object is further analyzed and attributes are assigned to the node object depending upon what rules, if any, have been defined for the currently available input. As an example, if the colour *red* is detected on a *card*, then a node representing *red* is added. An edge going from *card* to *red*, that is, from governant to dependant, is then created labeled with *adjectivial modifier*. Once no more classifiable attributes can be found, the VDPT, representing the object, can be compared with the DPT of the utterance describing the object. We define similarity as the percentage of identical triples in the two trees.

A visual representation of similarity between representations can be seen in Figure 1. Once similarity has been established, the robot has enough information to execute the command.

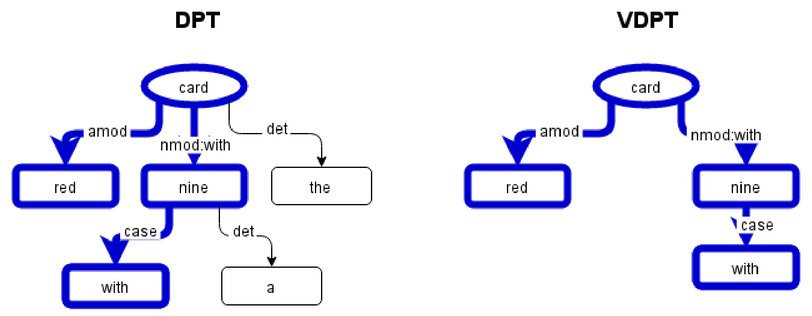

Figure 1: Example of a DPT of the sentence "the red card with a nine" and a VDPT of a perceived object (i.e. a red card with a nine on it).

References

[Bugmann et al.2005] G. Bugmann, JC. Wolf, P. Robinson. 2005. The impact of spoken interfaces on the design of service robots. In *Industrial Robot: An International Journal*, 32(6):499–504. Emerald Group Publishing Limited.

[Scheutz et al.2007] M. Scheutz, P. Schermerhorn, J. Kramer, D. Anderson. 2007. First steps toward natural human-like HRI. In *Autonomous Robots*, 22(4):411–423. Springer.

[Das et al.2010] D. Das, N. Schneider, D. Chen, N. Smith. 2010. Probabilistic frame-semantic parsing. In *Human language technologies: The 2010 annual conference of the North American chapter of the association for computational linguistics*, 948–956. Association for Computational Linguistics.

[Marneffe et al.2008] D. Marneffe, M. Manning, C. Manning. 2008. Stanford typed dependencies manual. *Technical report*. Stanford University, 2008.

[Sutherland2016] A. Sutherland. 2016. Using Dependency Parse Trees as a Method for Grounding Verbal Descriptions to Perceived Objects, *Proceedings of Umeå's 20th student conference in computing science: USCCS 2016*, 95–106. Umeå universitet, 2016.

[Sutherland et al.2015] A. Sutherland and S. Bensch and T. Hellström. 2015. Inferring Robot Actions from Verbal Commands Using Shallow Semantic Parsing. *Proceedings on the International Conference on Artificial Intelligence 2015 (ICAI 2015)*, pages 28–34.