Generalizing Context for Effective Grounding of Ambiguous Language

Wil Thomason, wbthomason@cs.cornell.edu

1 Problem Definition

Context is critical to understanding human communication. It is used by humans to encode knowledge assumed to be shared, disambiguate between otherwise equivalent options, and simply convey things more efficiently than might otherwise be possible. Thus, in order for robotic teams to be able to coexist with humans and collaborate on complex tasks, we must have a general means of understanding contextual information and including it in the language grounding process.

As detailed in the related work, the problem of context inclusion has previously been studied with limited success. Most approaches are only capable of incorporating very limited forms of context (i.e. basic knowledge about the environment configuration), require retraining of the model for changes in context, and/or insufficiently capture the temporally varying nature of context.

We think that we can succeed in this area where others have failed due to two factors: First, we can use insights into the nature of context specifically in terms of its use for grounding to create a sufficiently abstract mathematical framework for context to describe general sources. Second, the advent of deep learning approaches (specifically such things as sequence prediction and attention mechanisms) may prove useful for selecting salient context.

1.1 Goals

We want a system that can (a) select relevant context at any given point in an interaction and (b) incorporate general sources of context into the symbol grounding process to mitigate potential ambiguous/mistaken groundings.

By (a) we mean that the system should be able to take in a record of an interaction to a certain point in time and determine a weighting of a set of sources of context which corresponds to how much they matter at the current time[1].

[1] This task seems remarkably similar to the context construction problem. Maybe there's a higher level/more general problem of which they're both instances that we could use to solve them both at once?

By (b) we mean that the system should be able to take an abstract representation of a source of context adhering to some predefined interface and use it to correctly ground symbols in cases where the result would be ambiguous or incorrect without knowledge of the context. Sources of context we would like to be able to encode (as examples) include minimally:

Environmental Context: Knowledge of concrete objects and actions, as well as properties of the same.

Historical Context: Essentially, memory of what has been important thus far in the interaction/memory of the interaction thus far.

Task Context: Semantic understanding of the current task (and subtasks, etc.), which informs what actions and objects are likely to be salient.

Physical Feasibility: Introspective understanding of the physical properties of the robot and of objects in the environment. This includes things like reachability, ability to manipulate in a certain way, etc.

Non-Verbal Entity Indicators: Gestures, etc. that can indicate a certain action or object.

2 Related Work

- Misra et al. (2016) use an energy function approach to grounding to encode limited environmental context. They can't capture things like historical context, general object properties (the model they learn can do object similarity matching, but not things like **FINISH**)

3 Approach

3.1 Distribution-Modifying Functions

Intuitively, a context makes certain real-world entities more or less likely to be the correct groundings for certain abstract entities. This is supported by **SOURCES IN PRAGMATICS AND MAYBE PSYCH?**.

As such, we model a generalized context as a **probability distribution-modifying function**. That is, a context can be considered[2] as follows:

Definition 1. For $\mathbb{P}$ the set of all probability distributions, a **context** is a function $c : \mathbb{P} \to \mathbb{P}$ such that for distributions $p, p' \in \mathbb{P}$ and such that $c(p) = p'$,

[2]From the point of view of the grounding operation, at least - there's obviously some associated data and computation included in a context outside of this.

TODO: The problem here is that I don't know how to construct instances of the function interface, and what I can construct - the voting algorithm - doesn't fit with the intuition or interface or feel novel enough.

3.2 Selecting Context

This is a later stage of the project. Loosely, I have some thoughts regarding adapting attention mechanisms from neural networks, but I need to give this more thought.

4 Evaluation

References

Misra, Dipendra K, Jaeyong Sung, Kevin Lee, and Ashutosh Saxena. 2016. "Tell Me Dave: Context-Sensitive Grounding of Natural Language to Manipulation Instructions." *The International Journal of Robotics Research* 35 (1-3): 281–300.