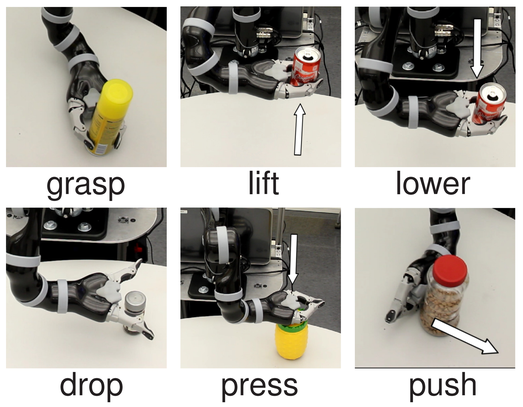
Guiding Interaction Behaviors for Multi-modal Grounded Language Learning

Jesse Thomason, Jivko Sinapov, Raymond J. Mooney
University of Texas at Austin

Multi-Modal Grounded Linguistic Semantics

Robots need to be able to connect language to their environment in order to discuss real world objects with humans. *Grounded language learning* connects a robot's sensory perception to natural language predicates. We consider a corpus of objects explored by a robot in previous work (Sinapov, IJCAI 2016).



A *hold* and *look* behavior were also performed for each object.

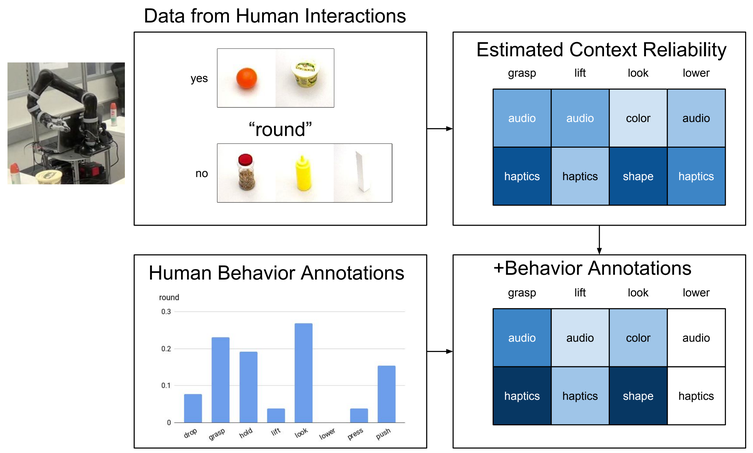
- Objects sparsely annotated with language predicates from an interactive "I Spy" game with humans (Thomason, IJCAI 2016).
- Many object examples are necessary to ground language in multi-modal space
- We explore additional sources of information from users and corpora
  - Behavior annotations, "What behavior would you use to understand 'red'?"
  - Modality annotations, "How do you experience 'heaviness'?"
  - Word embeddings, exploiting that "thin" is close to "narrow."

Methodology

The decision $d(p, o) \in [-1, 1]$ for predicate p and object o is

$$d(p, o) = \sum_{c \in C} \kappa_{p,c} G_{p,c}(o), \quad (1)$$

for $G_{p,c}$ a linear SVM trained on labeled objects for p in the feature space of context c with Cohen's $\kappa_{p,c}$ agreement with human labels during cross-validation.

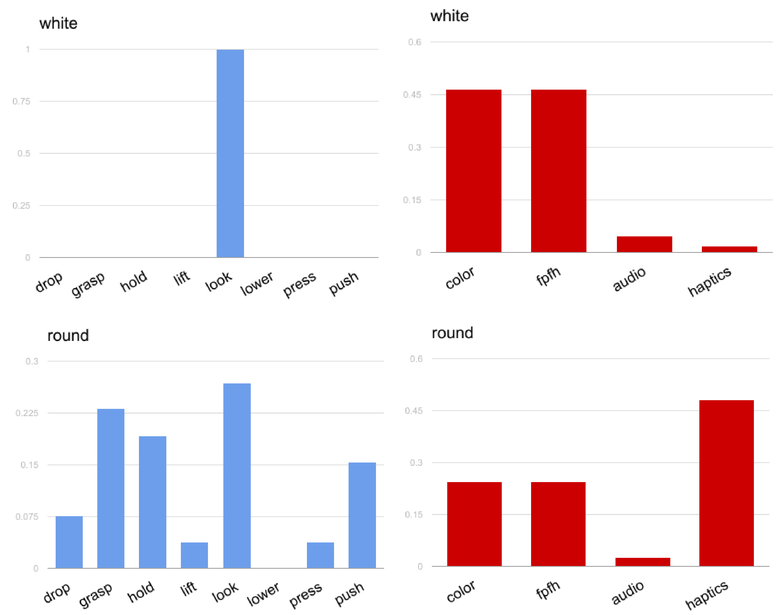


Behavior Annotations. We gather relevant behaviors from human annotators for each predicate, allowing us to augment classifier confidences with human suggestions. The predicate "heavy" favors interaction behaviors like lifting and holding.

Modality Annotations. We also derive a modality annotations past work (Lynott, 2009), allowing us to augment classifier confidences with human intuitions. The predicate "white" favors visual-related modalities.

Word Embeddings. We calculate positive similarity as and subsequently get augment each predicate's classifier confidences with similarity-weighted averages of its predicate neighbors. A predicate like "narrow" with few object examples can borrow information from a predicate like "thin" that is close in word embedding space and has many examples.

Experiment and Results



	p	r	f1
mc	.282	.355	.311
κ	.406	.460	.422
B+κ	.489	.489	.465
M+κ	.414	.466	.430
W+κ	.373	.474	.412

- Adding behavior annotations or modality annotations improves performance over using kappa confidence alone.
- Behavior annotations helped the *f*-measure of predicates like "pink", "green", and "half-full"
- Modality annotations helped with predicates like "round", "white", and "empty".
- Sharing kappa confidences across similar predicates based on their embedding cosine similarity improves recall at the cost of precision.
- For example, "round" improved, but at the expense of domain-specific meanings of predicates like "water".

jesse@cs.utexas.edu https://www.cs.utexas.edu/~jesse/