Grounded language understanding: Neural RSA

Christopher Potts

Stanford Linguistics

CS224u: Natural language understanding
Papers employing these techniques

- Andreas and Klein 2016
- Fried et al. 2018
- Monroe et al. 2017
- Monroe et al. 2018
Motivation

- Discriminative image labeling
- Image captioning
- Machine translation
- Collaborative problem solving
- Interpreting complex descriptions
- Optical Character Recognition

- Scalability
- Sensitivity to variation
- Bounded rationality
- New kinds of model assessment
- Impact
# Colors in context

<table>
<thead>
<tr>
<th>Context</th>
<th>Utterance</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="blue.png" alt="Blue" /></td>
<td>blue</td>
</tr>
<tr>
<td><img src="darker_blue.png" alt="Darker Blue" /></td>
<td>The darker blue one</td>
</tr>
<tr>
<td><img src="teal.png" alt="Teal" /></td>
<td>teal not the two that are more green</td>
</tr>
<tr>
<td><img src="dull_pink.png" alt="Dull Pink" /></td>
<td>dull pink not the super bright one</td>
</tr>
<tr>
<td><img src="not_green.png" alt="Not Green" /></td>
<td>not any of the regular greens</td>
</tr>
<tr>
<td><img src="purple.png" alt="Purple" /></td>
<td>Purple</td>
</tr>
<tr>
<td><img src="blue.png" alt="Blue" /></td>
<td>blue</td>
</tr>
</tbody>
</table>

Stanford Colors in Context corpus (Monroe et al. 2017)
Literal neural speaker $s_{\text{lit}}^\theta$
Neural literal listener

Encoder

Decoder

score(f_i) = -(f_i - μ)^TΣ(f_i - μ)

Fourier transform

softmax(s_1, s_2, s_3)

Monroe et al. 2017
Neural pragmatic agents

Neural pragmatic speaker (Andreas and Klein 2016)

\[
S_{\text{prag}}^\theta(msg \mid state) = \frac{L_0^\theta(state \mid msg)}{\sum_{msg' \in X} L_0^\theta(state \mid msg')}
\]

with \(X\) a sample from \(S_{\text{lit}}^\theta(msg \mid state)\) such that \(msg \in X\).

Neural pragmatic listener

\[
L_1^\theta(state \mid msg) \propto S_{\text{prag}}^\theta(msg \mid state)
\]

Blended neural pragmatic listener

Weighted combination of \(L_0^\theta\) and \(L_1^\theta\).
Other related work

- Golland et al. (2010): Recursive speaker/listener reasoning as part of interpreting complex utterances compositionally, with grounding in a simple visual world.
- Wang et al. (2016): Pragmatic reasoning helps in online learning of semantic parsers.
- Tellex et al.’s (2014) Inverse Semantics: Robot utterances are scored by models similar to RSA’s pragmatic speakers.
- Cohn-Gordon and Goodman (2019): RSA for translation
- Monroe and Potts (2015): “RSA as a hidden activation function”
- Mao et al. 2016: pragmatic learning objectives


