Abstract
- Humans adhere to natural language because they have to interact with an entire community
- Having a private language for each person would be inefficient

We propose a multi-agent dialog framework (MADF) where each agent interacts with and learns from multiple agents and show that it results in more coherent and human-interpretable dialog between agents, without compromising on task performance.

Problem Statement
- Formulated as a conversation between two collaborative agents, a Question (Q-) Bot and an Answer (A-) Bot
- A-Bot given an image, while Q-Bot is given only a caption to the image - both agents share a common objective, which is for Q-Bot to form an accurate mental representation of the unseen image
- Facilitated by exchange of 10 pairs of questions and answers between the two agents, using a shared common vocabulary
- Pretraining the agents with supervision from the VisDial dataset, followed by making them interact and adapt to each other via reinforcement learning maximizes task performance, but the agents learn to communicate in non-grammatical and semantically meaningless sentences, hence motivating our multi-agent setup.

Method
- 1 Q-Bot and 1 A-Bot are trained in isolation via supervision to optimize the MLE objective, leading to uninformative and repetitive dialog and inability to respond to out of distribution questions/answers
- We use Curriculum based learning to smoothly transition from supervised learning to Reinforcement Learning
- Reinforcement Learning uses the change in distance between the predicted image embedding and the ground truth embedding as the reward function which is shared by both the Q and A bot. We train the system using the REINFORCE algorithm.
  \[
  r_t = (y_t^2 - \langle y_t, y_t^* \rangle) - \langle y_t, y_t^* \rangle
  \]
  - No explicit incentive to maintain natural language and hence prone to deviate from it to optimize transfer of information between bots
- We solve the problem using our Multi Agent setup where we arbitrarily pick a Q and A bot pair and carry out a round of training for them and keep repeating the process
- Much harder for the bots to deviate from natural language in this setting as coming up with a new language pair for each pair of bots is highly inefficient

Algorithm 3 Multi-Agent Dialog Framework (MADF)

1. procedure MADF \( Q\) \( A \) \( Q\) \( A \)
2. \( Q \leftarrow \text{random select} \) \( Q \), \( Q \), \( Q \)
3. \( A \leftarrow \text{random select} \) \( A \), \( A \), \( A \)
4. \( Q \leftarrow \text{fixed} \) \( Q \)
5. \( A \leftarrow \text{fixed} \) \( A \)
6. \( \text{end for} \)
7. \( \text{end procedure} \)

Table 1: Comparisons of Metrics with Literature

<table>
<thead>
<tr>
<th>Model</th>
<th>MRR</th>
<th>Rank Mean</th>
<th>R1@5</th>
<th>R5@5</th>
<th>R10@5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Answer Prior (Das et al., 2018)</td>
<td>0.3725</td>
<td>26.90</td>
<td>23.95</td>
<td>48.52</td>
<td>53.23</td>
</tr>
<tr>
<td>MN- Quad (Das et al., 2016)</td>
<td>0.5259</td>
<td>17.06</td>
<td>42.79</td>
<td>68.88</td>
<td></td>
</tr>
<tr>
<td>HC-EG-GDI (Li et al., 2017)</td>
<td>0.547</td>
<td>14.23</td>
<td>44.35</td>
<td>65.78</td>
<td>71.55</td>
</tr>
<tr>
<td>EntityQ-Multi (Das et al., 2017)</td>
<td>0.457</td>
<td>21.11</td>
<td>50.87</td>
<td>60.48</td>
<td></td>
</tr>
<tr>
<td>CoSaGAN (Wu et al., 2017)</td>
<td>0.5758</td>
<td>14.44</td>
<td>46.10</td>
<td>65.59</td>
<td>71.74</td>
</tr>
<tr>
<td>SL(Den)</td>
<td>0.619</td>
<td>3.223</td>
<td>34.74</td>
<td>56.77</td>
<td>72.68</td>
</tr>
<tr>
<td>DL (Q1, A1) (Den)</td>
<td>0.659</td>
<td>7.007</td>
<td>16.64</td>
<td>54.60</td>
<td>72.34</td>
</tr>
<tr>
<td>DL (Q2, A2) (Den)</td>
<td>0.601</td>
<td>5.495</td>
<td>34.83</td>
<td>57.47</td>
<td>72.48</td>
</tr>
<tr>
<td>DL (Q3, A3) (Den)</td>
<td>0.590</td>
<td>5.56</td>
<td>33.50</td>
<td>57.37</td>
<td>72.61</td>
</tr>
</tbody>
</table>

Figure 1: Multi-Agent (with 1 Q-Bot, 3 A-Bots) Dialog Framework

Figure 2: Two randomly selected images from the VisDial dataset followed by the ground truth (human) and generated dialog about that image for each of 4 systems (Q4, R4, Q1, R1, Q2, R2, Q3, R3).

Future Work
- We plan to explore several other multi bot architectural settings and perform a more thorough human evaluation for qualitative analysis of our dialog.
- We also plan on incorporating other language priors in our reinforcement learning setup to further improve the dialog quality.
- We will also experiment with using a discriminative answer decoder which uses information of the possible answer candidates to rank the generated answer with respect to all the candidate answers and use the ranking performance to train the answer decoder.

References