The Task: Generate coherent and controlled responses that follow system level goals

Motivation

Dialogue systems based on large LMs (DialoGPT[1], Blenderbot[2], etc.)
- Coherent responses
- Follow system-level goals
- Require labelled data for each new goal

Dialogue systems based on control based on persona, emotions, knowledge etc.
- Coherent responses
- Follow system-level goals
- Use control signals

EDGE's key advantages are its Gold response Tokens

Exemplar based control

Dialogue Context

Exemplar Response

Exemplar-Driven Response

Prior exemplar models
Use Exemplar Tokens?

Ignores exemplar

Copies irrelevant tokens

Don't have you been always nervous about flying in the past?

Qualitative Results

(1) EDGE > all other models
(2) frame-based models > token-based models
(3) GPT-2-based models > LSTM-based models

Quantitative Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Coherence</th>
<th>Fluent</th>
<th>Interesting</th>
<th>Engagement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dialog</td>
<td>0.241</td>
<td>2.61</td>
<td>2.48</td>
<td>2.32</td>
</tr>
<tr>
<td>GPT2-Geo</td>
<td>0.249</td>
<td>2.55</td>
<td>2.41</td>
<td>2.34</td>
</tr>
<tr>
<td>LSTM-Geo</td>
<td>0.182</td>
<td>2.30</td>
<td>2.19</td>
<td>2.24</td>
</tr>
<tr>
<td>EDGE</td>
<td>0.278</td>
<td>2.63</td>
<td>2.56</td>
<td>2.39</td>
</tr>
<tr>
<td>Human</td>
<td>0.383</td>
<td>2.70</td>
<td>2.78</td>
<td>2.44</td>
</tr>
</tbody>
</table>

Advantages of EDGE

- Coherent responses
- Follow system-level goals
- No explicit training labels

Experiment: Open-domain

Dataset: DailyDialog[3]

Experiment: Scam Defense

Dataset: Scam Email Defense[4]

Experiment: Open-domain

Dialogue Context

Exemplar Response

Exemplar Driven Response

Prior exemplar models
Use Exemplar Tokens?

Ignores exemplar

Copies irrelevant tokens

Don't have you been always nervous about flying in the past?

Qualitative Results

(1) EDGE generates longer and more specific responses
(2) EDGE generates coherent responses even with irrelevant or missing frames
(3) EDGE occasionally diverges from exemplar response

Quantitative Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Coherence</th>
<th>Fluent</th>
<th>Interesting</th>
<th>Engagement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dialog</td>
<td>0.241</td>
<td>2.61</td>
<td>2.48</td>
<td>2.32</td>
</tr>
<tr>
<td>GPT2-Geo</td>
<td>0.249</td>
<td>2.55</td>
<td>2.41</td>
<td>2.34</td>
</tr>
<tr>
<td>LSTM-Geo</td>
<td>0.182</td>
<td>2.30</td>
<td>2.19</td>
<td>2.24</td>
</tr>
<tr>
<td>EDGE</td>
<td>0.278</td>
<td>2.63</td>
<td>2.56</td>
<td>2.39</td>
</tr>
<tr>
<td>Human</td>
<td>0.383</td>
<td>2.70</td>
<td>2.78</td>
<td>2.44</td>
</tr>
</tbody>
</table>

Advantages of EDGE

- Coherent responses
- Follow system-level goals
- No explicit training labels

Conclusion

- EDGE generates responses based on semantic frames of exemplar responses.
- EDGE achieves coherent responses that preserve system-level goals (implicitly present in exemplars) as demonstrated in our experiments.
- EDGE's key advantages are its controllability and zero-shot performance

Check out the code for this project!

https://github.com/prakhargupta/EDGE-exemplars

References


https://www.kaggle.com/rtatman/fraudulent

https://github.com/prakhargupta/EDGE-exemplars