Argument mining, a growing field in natural language generation, includes the automatic identification and generation of argumentative structures within conversation.

We experiment with various methods for creating a dialogue agent that can engage in argumentative discourse.

Overview

- Problem
  - Argument mining, a growing field in natural language generation, includes the automatic identification and generation of argumentative structures within conversation.
  - We experiment with various methods for creating a dialogue agent that can engage in argumentative discourse.

- Significance
  - Utility in education and assessment as well as business use for investment decision.
  - Advances self-attention/transformer in argument NLG/NLU objectives.

Existing Approaches

- Current state-of-the-art generative model: hierarchical recurrent neural network, encoding and decoding at one level and updating a conversation-level state at another.
- Encoder: Bidirectional GRU encoder w/ conversation-level RNN memory.
- Decoder: Vanilla RNN.
- Model often misinterprets arguments or produces irrelevant responses.

Data

\[ d = [p(1), p(2), ..., p(m)] \]

\[ p(i) = [u(1)i, u(2)i, ..., u(n)i] \]

- Internet Argument Corpus Dataset-v1: 11,800 discussions w/ ~390,000 posts total.
- Training instance: discussion, \( d \) (sequence of posts).
- Gold instances are offset from train instances.
- \( p \) is a padded sequence of tokens, \( w \).

Approach

**Project Phases**

1. LSTM Seq2Seq - model baseline, context-free argument generation.
2. Pure transformers - context-free argument generation.
3. Transformer with LSTM Session Memory - context-rich argument generation.

**Approach**

- **Task**
  - Given a post (w, w/o context), generate an appropriate adversarial argumentative response.

**Additional Tunings**

1. Hyperparameter search - layers, dimensions, attention heads, learning rate, vocabulary size, min word count, etc.
2. Pre-training with cross-argumentative embedding objective (Self-referential).
3. GloVe embeddings vs. training from scratch.
4. <unk> thresholding, vocabulary pruning, etc. (16k size).

Results

**LSTM Seq2Seq (w/o char decoder)**

<table>
<thead>
<tr>
<th>Query</th>
<th>Output</th>
</tr>
</thead>
</table>
| I think the bible is real and it should be respected | I'm not sure that you are not a christian. I do think it is a matter of the bible. I do not think it is a matter of a person.
| Woman’s rights and suffrage should be upheld under the law | I do not know what I said, but I do not know what you are talking about.

**Transformer w/ LSTM**

<table>
<thead>
<tr>
<th>Query</th>
<th>Output</th>
</tr>
</thead>
</table>
| I believe in god | I'm not sure what you mean by "god". well then, have you heard of the bible? I do not think it's a good idea.
| I think that gun control has the potential to solve a lot of the problems with school shootings | I think that a good point. I think that if illegal because its effectively murder you are a christian, you are a christian.

**Table 1. Transformer w/ LSTM validation metrics with tuned parameters**

<table>
<thead>
<tr>
<th>Epoch</th>
<th>Perplexity</th>
<th>Accuracy</th>
<th>Loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>205.88</td>
<td>17.30%</td>
<td>5.33</td>
</tr>
<tr>
<td>4</td>
<td>84.21</td>
<td>24.83%</td>
<td>4.43</td>
</tr>
<tr>
<td>8</td>
<td>72.10</td>
<td>26.35%</td>
<td>4.28</td>
</tr>
<tr>
<td>12</td>
<td>65.65</td>
<td>27.36%</td>
<td>4.18</td>
</tr>
<tr>
<td>16</td>
<td>62.40</td>
<td>28.00%</td>
<td>4.13</td>
</tr>
</tbody>
</table>

Analysis

- From our qualitative results, we conclude that our dataset is ill-suited for generating more sophisticated language models typical of advanced argumentative discourse.
- Our extensive hyperparameter search suggests that our cross entropy training objective is overly simplistic for more complex generation tasks. A more involved theoretical formulation of training loss could yield qualitative translation improvements.
- We were impressed by the model’s ability to infer the underlying basis of the human input arguments.
- Additionally, the dialogue agent was proficient in establishing a sufficiently resolute position on many topics.

Future Work

- Less primitive argumentation datasets increases language model expressivity.
- Fine-tuning on pretrained contextual embeddings (BERT) captures word relationships more precisely for better NLG.
- More sophisticated attention mechanisms may allow for a more informative signal for decoding.

References

