Exemplar Encoder Decoder for Neural Conversation Generation

By

Gaurav Pandey, Danish Contractor, Vineet Kumar and Sachindra Joshi

IBM Research AI
Generative Models for Conversations

- **Context encoder**: (1) RNN (2) hierarchical RNN
- **Decoder**: RNN
- **Objective**: log probability of GT response given context.
- **Can generate novel responses for novel contexts!!**
Retrieval Models for Conversations

• Retrieve a response from a nearest neighbor index constructed from the training data.
• Can be used for closed domain problems.

• **Advantages:**
  • Answers are grounded in the domain.
  • Easy to prune answers according to requirements.

• **Disadvantage:**
  • Can not generate novel responses.

Can we use generative models to fix this?
Exemplar Encoder Decoder

• Build an index from all context-response pairs offline.
• For each context $c$:
  • Retrieve a set of exemplar contexts and corresponding responses.
  • Match the exemplar contexts with $c$ and get the similarities.
  • Use these similarities to weigh the exemplar responses.
Matching Exemplar Contexts

The normalized similarities are used to weigh the exemplar responses.
\[ ll = \log \sum_{k=1}^{K} s^{(k)} p(r | e^{(k)}) \]
Analyzing the Objective

Think of exemplar contexts and responses as latent variables

\[
\log p(r | c) = \log \sum_{(c', r')} p(r | c, r') \ p(c' | c) \\
\leq \log \sum_{1 \leq k \leq K} p(r | c, r^k) \ p(c^k | c) \\
= \log \sum_{1 \leq k \leq K} p(r | e^{(k)}) \ s^{(k)}
\]
Evaluation

• Exemplar Encoder Decoder
  • Hierarchical Recurrent Encoder
  • TF-IDF for retrieving exemplar conversations

• Datasets used:
  • Ubuntu Dialogue Corpus
  • IBM Tech Support Dataset

• Comparison Metrics
  • Activity and Entity metrics
  • Embedding metrics
Activity and Entity metrics

These metrics compare the precision, recall and F1 score of specific nouns and verbs present in the generated response as compared to the groundtruth response.

<table>
<thead>
<tr>
<th></th>
<th>Activity</th>
<th>Entity</th>
<th>Tense</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>R</td>
<td>F1</td>
<td></td>
</tr>
<tr>
<td><strong>Model</strong></td>
<td><strong>P</strong></td>
<td><strong>R</strong></td>
<td><strong>F1</strong></td>
<td><strong>P</strong></td>
</tr>
<tr>
<td>LSTM*</td>
<td>1.7</td>
<td>1.03</td>
<td>1.18</td>
<td>1.18</td>
</tr>
<tr>
<td>VHRED*</td>
<td>6.43</td>
<td>4.31</td>
<td>4.63</td>
<td>3.28</td>
</tr>
<tr>
<td>HRED*</td>
<td>5.93</td>
<td>4.05</td>
<td>4.34</td>
<td>2.81</td>
</tr>
<tr>
<td>EED</td>
<td>6.42</td>
<td>4.77</td>
<td>4.87</td>
<td>3.8</td>
</tr>
</tbody>
</table>

For comparison, the retrieval only model has an activity F1 score of **4.23** and entity F1 score of **2.72** respectively.
Embedding metrics

- These metrics compare the word embeddings of the generated response with the words of the groundtruth response.
- These metrics do not correlate with human judgements for Ubuntu Corpus\(^1\).

<table>
<thead>
<tr>
<th>Model</th>
<th>Ubuntu Average</th>
<th>Greedy</th>
<th>Extrema</th>
<th>Tech Support Average</th>
<th>Greedy</th>
<th>Extrema</th>
</tr>
</thead>
<tbody>
<tr>
<td>HRED</td>
<td>0.5131</td>
<td>0.4103</td>
<td>0.3263</td>
<td>0.6372</td>
<td>0.5100</td>
<td>0.4355</td>
</tr>
<tr>
<td>EED</td>
<td>0.5528</td>
<td>0.4172</td>
<td>0.3134</td>
<td>0.6420</td>
<td>0.5125</td>
<td>0.4356</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Input Context (C) and exemplar response (R)</th>
<th>Generated/Ground Truth Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1</strong>&lt;br&gt;C: i want to check ntfs file systems. is fsck ok with ntfs. → no , you ’d need to use chkdsk from windows. → R: i already symlinked ntfsfix to fsck. ntfs and am scanning the unmounted devices</td>
<td>HRED: i know , but i dont know how to do that&lt;br&gt;VHRED: how can i do that ?&lt;br&gt;EED: i am trying to use ntfs 3g i have a windows partition.&lt;br&gt;GT: i dont have windows installed</td>
</tr>
<tr>
<td><strong>2</strong>&lt;br&gt;C: anyone help me ? → seems that its detecting grub , have you tried checking continue without installing grub ?&lt;br&gt;R: try the live cd... you ca n’t boot into linux unless grub install</td>
<td>HRED: yes i did , but i have no idea how to fix it&lt;br&gt;VHRED: yes i did. i dont know how to do it&lt;br&gt;EED: nope not yet. grub install cd is not working&lt;br&gt;GT: no i did not.</td>
</tr>
<tr>
<td><strong>3</strong>&lt;br&gt;C: any ideas on my mplayer problem ? → reinstall. apt-get purge mplayer then reinstall&lt;br&gt;R: but half my system depends on this package ...</td>
<td>HRED: I did that , but it didn’t work&lt;br&gt;VHRED: I dont know how to do that&lt;br&gt;EED: does n t work. package does n t work&lt;br&gt;GT: that does n’t seem as a solution</td>
</tr>
</tbody>
</table>
Discussion

• A generative model that utilizes similar conversations for response generation.
• Can generate novel responses while ensuring that the responses are grounded in the domain.
• Incorporating retrieved conversations during generation improves performance as evident from several metrics.
• The proposed idea is general and can be used for image captioning and neural machine translation.