**Passage-level QG**

**Question:** How to do automatic natural question generation (QG) at document-level?

**Example:** From Wikipedia article Fresno, California

Q: Which is the largest city not connected to an interstate highway? A: Fresno

Q: Which State Route has been in discussion to upgrade to interstate standards? A: SR 99

Q: What are the factors that are contributing to the desire to have SR 99 improved to be of interstate standards? A: rapidly rising population and traffic in cities along SR 99, as well as the inability of Federal funding to meet the needs of the region.

**How2:** A first step, important (question-worthy) sentences selection.

**Why2:**
- Educational apps
- Question Answering
- Generating FAQs
- Conversational agent

**Hierarchical sequence tagging model**

**Task Objective:**
- **Input:** a paragraph D consisting of a sequence of sentences \( \{s_1, \ldots, s_m\}\)
- **Objective:** To select a subset of k question-worthy sentences \( (k < m)\), such that:

\[
P(y|D) = \text{softmax}(\text{MLP}(\text{enc}(s_1, \ldots, s_m)))
\]

where MLP is multi-layer neural network and tanh is the activation function.

The first layer encodes the sentences of the paragraph, with two encoding alternatives:
- CNN + max pooling
- Sum operation

**Sentence-level question generation**

- Encoding only sentence as input, do not consider paragraph/context-level, attending to source sentence hidden states.

**Automatic Evaluation**

<table>
<thead>
<tr>
<th>Model</th>
<th>Precision</th>
<th>Recall</th>
<th>F-measure</th>
<th>Acc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>RANDOM</td>
<td>83.45</td>
<td>50.20</td>
<td>56.11</td>
<td>50.27</td>
</tr>
<tr>
<td>Majority Baseline</td>
<td>63.21</td>
<td>100.00</td>
<td>77.46</td>
<td>63.21</td>
</tr>
<tr>
<td>CNN (Kim, 2014)</td>
<td>68.35</td>
<td>90.13</td>
<td>77.74</td>
<td>67.38</td>
</tr>
<tr>
<td>LSTM (Lindberg et al., 2015)</td>
<td>76.52</td>
<td>65.53</td>
<td>65.37</td>
<td>76.52</td>
</tr>
<tr>
<td>LSTM (Lindberg et al., 2015)</td>
<td>90.49</td>
<td>89.08</td>
<td>78.70</td>
<td>90.49</td>
</tr>
</tbody>
</table>

**Experiments**

- Our hierarchical sequence-tagging model beats the strong linear system.
- Majority forms a very strong baseline.
- Pre-trained word embeddings do not help significantly.

**Full QG Evaluation**

<table>
<thead>
<tr>
<th>Metric</th>
<th>BLEU 1</th>
<th>BLEU 2</th>
<th>BLEU 3</th>
<th>BLEU 4</th>
<th>METEOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conservative</td>
<td>70.90</td>
<td>63.90</td>
<td>90.13</td>
<td>90.13</td>
<td>90.13</td>
</tr>
<tr>
<td>Danau et al.</td>
<td>73.02</td>
<td>69.21</td>
<td>80.32</td>
<td>80.32</td>
<td>80.32</td>
</tr>
<tr>
<td>Parveen et al.</td>
<td>73.15</td>
<td>69.20</td>
<td>80.42</td>
<td>80.42</td>
<td>80.42</td>
</tr>
<tr>
<td>Our method</td>
<td>74.35</td>
<td>66.11</td>
<td>79.60</td>
<td>79.60</td>
<td>79.60</td>
</tr>
</tbody>
</table>

**Open questions:**
- Connecting QA and QA!
- Better dataset for the task.

**Conclusion**

- We introduce the new task of question-worthy sentences selection.
- We introduce a neural sentence-level sequence tagging approach for this task, with sum or CNN for encoding the sentences.
- Our system outperforms the baselines (e.g., feature-rich linear systems) significantly.