Interactive Neural Machine Translation

Sebastin Santy\textsuperscript{1}  Sandipan Dandapat\textsuperscript{2}  Monojit Choudhury\textsuperscript{1}  Kalika Bali\textsuperscript{1}

\textsuperscript{1}Microsoft Research, Bangalore, India  \textsuperscript{2}Microsoft R&D, Hyderabad, India
\{t-sesan, sadandap, monojitc, kalikab\}@microsoft.com

Aim

Assist translators by providing translation suggestions on the fly.

Advantages

- Faster Turnaround
  The gisting and suggestions help the translator breeze through translation tasks with minimal typing.
- High Translation Quality
  Language is inherently divergent and human translators cannot quickly enumerate all acceptable variants of a translation. On the other hand, machine translation has not yet reached human quality, though it can provide a number of variants. Combine the individual strengths, to produce high quality translations.
- Amateur Translators
  Expert translators are scarce. Take help of bilingual speakers who have native proficiency in these languages for translation tasks by providing suggestions and gisting.

Method

- **Seq2Seq decoder**: Conditional probability of generating output token $y_t$ at time step $t$, given the full input sequence $x$ and the previously output tokens $y_1, ..., y_{t-1}$ is:
  \begin{equation}
  p(y_t|y_1, ..., y_{t-1}, x) = g(y_{t-1}, s_t, c_t)
  \end{equation}
  - $g$ → non-linearity function
  - $s_t$ → hidden state
  - $c_t$ → context vector: weighted average of all encoder hidden states with weights generated by the attention mechanism

- **INMT decoder**: Condition based on the partial input from the human translator $\{y'_1, ..., y'_{t-1}\}$ instead of default Seq2Seq output $\{y_1, ..., y_{t-1}\}$:
  \begin{equation}
  p(y'_t|y'_1, ..., y'_{t-1}, x) = g(y'_{t-1}, s_t, c_t)
  \end{equation}

- **Sparsemax Attention** is used to aid one-to-one source-target word mapping for word coverage visualization.
- **Beam Search** is used to produce multiple suggestions based on the partial input. It selects the most probable full translation for a given input sentence. If and when the translator diverges from this full translation, a new beam search is conducted from the partial input prefix till end of sentence is encountered.

Interface Overview

- **Translation Gisting**
  Prime the translator with a quick translation to reduce cognitive load. Spotting errors in the gisting is much easier than trying to mentally structure the translations.
- **Translation Suggestions**
  Gist might not be the correct translation. Provide bi-gram suggestions which the translator can choose instead of the full gist.
- **Word Coverage Visualization**
  Show one-to-one source-target word mapping. This will help in understanding how much translation is completed.
- **Transliteration**
  Non-European languages have non-Latin script. Amateur translators usually use English keyboards to type. Provide character-wise transliteration, as each character triggers the engine to give new outputs.

Experiments

- **BLEU %** - Measure the BLEU score of the generated gist after a certain fraction - $x\%$ of words of the intended translation has been provided. Table 1 shows the average BLEU score for each language pair at different values of $x$.

<table>
<thead>
<tr>
<th>Data Size</th>
<th>0%</th>
<th>10%</th>
<th>20%</th>
<th>40%</th>
</tr>
</thead>
<tbody>
<tr>
<td>bn-en</td>
<td>1.1M</td>
<td>25.31</td>
<td>27.54</td>
<td>35.68</td>
</tr>
<tr>
<td>hi-en</td>
<td>1.5M</td>
<td>40.64</td>
<td>42.06</td>
<td>47.90</td>
</tr>
<tr>
<td>ml-en</td>
<td>897K</td>
<td>19.76</td>
<td>21.95</td>
<td>29.84</td>
</tr>
<tr>
<td>ta-en</td>
<td>428K</td>
<td>18.71</td>
<td>20.90</td>
<td>27.05</td>
</tr>
<tr>
<td>te-en</td>
<td>104K</td>
<td>11.92</td>
<td>14.57</td>
<td>21.17</td>
</tr>
</tbody>
</table>

- **Keystroke Reduction** - Algorithmically compare minimum number of keystrokes required when typing interactively versus the same when manually typing. Reduction of around 30% keystrokes is observed for all the above mentioned languages.

System Overview

OpenNMT (PyTorch); JQuery (JS); Django;

Try it out! https://aka.ms/inmt