Overview

Translation between distant language pairs are difficult
1. Some languages use many words to represent one thing while others use less words
2. Some languages are free word-order while others are not

Difficulties of translation of En -> Ja

Japanese sentence has
- longer sequence (En: 25 vs. Ja: 30 [words/sentence])
- free chunk order (e.g., 「だれかが / 犬に」 = 「犬に / だれかが」)

Proposal: Chunk-based Decoder for Neural Machine Translation

I heard that someone was bitten by a dog, weren’t you injured?

Two step decoding
- First a chunk, then words inside the chunk

Models fixed word order and free word order independently!

Translation between distant languages (En -> Ja)

Word-based Encoder-Decoder

Enc-Dec with Attention [Bahdanau+ 15]
- encodes / decodes “word-by-word”

Two additional connections
1. to capture the interaction between chunks
2. to memorize previous outputs well

Models fixed word order and free word order independently!

Proposal: Chunk-based Decoder for Neural Machine Translation

Two step decoding
- First a chunk, then words inside the chunk

Models fixed word order and free word order independently!

Experiments

Data
- ASPEC [Nakazawa+ 16], 1.6M En/Ja pairs

Preprocessing
- Bunsetsu chunking with J.DepP [Yoshinaga & Kitsuregawa 09]

Baseline systems
1. Word-based encoder-decoder [Bahdanau+ 15]
2. Tree-based encoder [Eriguchi+ 16] (SOTA)

Results

<table>
<thead>
<tr>
<th>Model</th>
<th>BLEU</th>
<th>RIBES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word-based encoder</td>
<td>37.26</td>
<td>82.23</td>
</tr>
<tr>
<td>+ Chunk-based decoder [Proposed]</td>
<td>35.81</td>
<td>81.29</td>
</tr>
<tr>
<td>Tree-based encoder [Eriguchi+ 16] + Word-based decoder</td>
<td>34.70</td>
<td>81.01</td>
</tr>
<tr>
<td>Word-based encoder-decoder [Bahdanau+ 15]</td>
<td>34.91</td>
<td>81.66</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Decoder type</th>
<th>BLEU</th>
<th>RIBES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chunk-based [Model 3]</td>
<td>8.69</td>
<td>52.82</td>
</tr>
<tr>
<td>Chunk-based [Model 2]</td>
<td>7.78</td>
<td>51.48</td>
</tr>
<tr>
<td>Chunk-based [Model 1]</td>
<td>7.59</td>
<td>50.47</td>
</tr>
<tr>
<td>Word-based</td>
<td>7.56</td>
<td>50.73</td>
</tr>
</tbody>
</table>

Quality of generated chunks

| Source | the atmospheric glow discharge is a homogeneous electric discharge obtained by applying alternating voltage after introducing atmospheric He gas in a typical dielectric barrier discharge reactor.
| Reference | 大気圧下で発生する等の高 |
| Word-based encoder-decoder [Bahdanau+ 15] | because user operation is | 回, 物質の操作が重要である |
| Chunk-based [Proposed] | important for the idea | もののためのインタフェ |

Scores inside () are not our implementations