Exploiting Cross-Sentence Context for Neural Machine Translation

Longyue Wang♥ Zhaopeng Tu♠ Andy Way♥ Qun Liu♥

♥ ADAPT Centre, Dublin City University ♠ Tencent AI Lab
Motivation

- The majority of NMT models is sentence-level
Motivation

- The continuous vector representation of a symbol encodes *multiple dimensions of similarity*.

<table>
<thead>
<tr>
<th>Word $x'$</th>
<th>Axis</th>
<th>Nearest Neighbours</th>
</tr>
</thead>
<tbody>
<tr>
<td>notebook</td>
<td>1</td>
<td>diary notebooks (notebook) sketchbook jottings</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>palmtop notebooks (notebook) ipaq laptop</td>
</tr>
<tr>
<td>power</td>
<td>1</td>
<td>powers authority (power) powerbase sovereignty</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>powers electrohydraulic microwatts hydel (power)</td>
</tr>
</tbody>
</table>

(Choi et al., 2016)
Motivation

- The continuous vector representation of a symbol encodes *multiple dimensions of similarity*.

- **Consistency** is another critical issue in document-level translation.

<table>
<thead>
<tr>
<th>Past</th>
<th>Current</th>
</tr>
</thead>
<tbody>
<tr>
<td>well, on this <em>issue</em>, Iran has a relatively …</td>
<td>that just mentioned the <em>issue</em> of the talks …</td>
</tr>
<tr>
<td>in <em>his</em> term, …</td>
<td>…</td>
</tr>
<tr>
<td>in <em>the</em> Iranian nuclear <em>issue</em> in his term, …</td>
<td>…</td>
</tr>
</tbody>
</table>
Motivation

• The **cross-sentence context** has proven helpful for the aforementioned two problems in multiple sequential tasks (Sordoni et al., 2015; Vinyals and Le, 2015; Serban et al., 2016).
Motivation

• The **cross-sentence context** has proven helpful for the aforementioned two problems in multiple sequential tasks (Sordoni et al., 2015; Vinyals and Le, 2015; Serban et al., 2016).

• However, it has *received relatively little attention* from the NMT research community.
Data and Setting

- Chinese-English translation task
- Training data: 1M sentence pairs from LDC corpora that contain document information
- Tuning: NIST MT05, Test: NIST MT06 and MT08
- Build the model on top of Nematus (https://github.com/EdinburghNLP/nematus)
- Vocabulary size: 35K for both languages
- Word embedding: 600; Hidden size: 1000
Approach

• Use a *Hierarchical RNN* to summarize previous M source sentences.

Word-level RNN

Sentence-level RNN

Cross-Sentence Context
Approach

- Strategy I: **Initialization** — *Encoder*

Cross-Sentence Context

This is an ecological network.
Approach

- Strategy I: **Initialization** — *Decoder*

![Diagram with symbols and arrows representing the flow of sentences and context](image-url)
Approach

• Strategy I: Initialization — Both
Results

- Impact of components
Approach

- Strategy 2: Auxiliary Context

Cross-Sentence Context

Intra-Sentence Context
Approach

• Strategy 2: Auxiliary Context

(a) standard decoder
(b) decoder with auxiliary context
(c) decoder with gating auxiliary context
Results

• Impact of components

- Baseline: 30.57
- +Aux. Ctx.: 31.3
- +Gating Aux. Ctx.: 32.24
Approach

- Initialization + Gating Auxiliary Context

Cross-Sentence Context

Intra-Sentence Context
Results

- Impact of components
Analysis

• Translation error statistics

<table>
<thead>
<tr>
<th>Errors</th>
<th>Ambiguity</th>
<th>Inconsistency</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>38</td>
<td>32</td>
<td>70</td>
</tr>
<tr>
<td>Fixed</td>
<td>29</td>
<td>24</td>
<td>53</td>
</tr>
<tr>
<td>New</td>
<td>7</td>
<td>8</td>
<td>15</td>
</tr>
</tbody>
</table>
Analysis

• Case Study

<table>
<thead>
<tr>
<th>Hist.</th>
<th>这 不 等于 明着 提前 告诉 贪官 们 赶紧 转移 罪证 吗？</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input</td>
<td>能否 遏制 和 震慑 腐官？</td>
</tr>
<tr>
<td>Ref.</td>
<td>Can it inhibit and deter corrupt officials?</td>
</tr>
<tr>
<td>NMT</td>
<td>Can we contain and deter the <em>enemy</em>?</td>
</tr>
<tr>
<td>Our</td>
<td>Can it contain and deter the <em>corrupt officials</em>?</td>
</tr>
</tbody>
</table>
Summary

• We propose to use HRNN to summary previous source sentences, which aims at providing cross-sentence context for NMT

• Limitations
  • Computational expensive
  • Only exploit source sentences due to error propagation
  •Encoded into a single fixed-length vector, not flexible
Publicly Available

• The source code is publicly available at https://github.com/tuzhaopeng/LC-NMT

• The trained models and translation results will be released
Reference


Question & Answer