Overview

Q: Can we learn to compose words into sentence representations using parallel text as supervision?
A: yes, and bilingual phrases outperform bilingual sentences and monolingual paraphrases

Background: learning sentence representations
- from PPDB paraphrases can outperform complex (un)supervised models (Paragram, Wieting et al’16)
- from bilingual sentence pairs can be useful for cross-lingual document classification (e.g. BiCVM, Hermann and Blunsom’14)

Models

- Composition by word-averaging (Wieting et al’16)
  \[ g(\text{piggy saw kermit dancing}) = \frac{1}{4} ( \text{piggy saw} + \text{kermit} + \text{dancing} ) \]
- Learn word representations using semantically equivalent text segment pairs \((x_1, x_2)\)

\[
\min \sum_{(x_1, x_2)} [\delta + \|g(x_1) - g(x_2)\|^2] \quad \text{minimize distance between equivalent examples}
\]
\[
- \|g(x_1) - g(x_2)\|^2 \quad \text{minimize distance between far apart examples}
\]

Evaluation Setup

2 sentence-level similarity tasks
- SICK (SemEval’14) (designed for evaluating composition models)

Evaluation metric: Pearson correlation of cosine similarity & gold scores

<table>
<thead>
<tr>
<th>Training Conditions</th>
<th>#pairs</th>
<th>Avg. Len</th>
<th>Provenance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bilingual sentences</td>
<td>1.9M</td>
<td>28</td>
<td>en-es europarl-v7</td>
</tr>
<tr>
<td>Bilingual phrases</td>
<td>3M</td>
<td>5</td>
<td>Moses phrase extraction</td>
</tr>
<tr>
<td>Monolingual phrases</td>
<td>3M</td>
<td>3</td>
<td>XL PPDB</td>
</tr>
</tbody>
</table>

Findings

1. Bilingual phrases outperform monolingual paraphrases and parallel sentences in controlled settings

2. Training with a compositional objective requires less data

3. Bilingual phrases avoid the lexical overlap issue in paraphrases

\[
\| (\text{healthy + and + stable} ) - (\text{healthy + and + steady} ) \| = \| \text{stable} - \text{steady} \|
\]

References


Karl Moritz Hermann and Phil Blunsom, “Multilingual models for compositional distributed semantics”. In ACL, 2014.